Focusing on the individual infant: classification and heterogeneity of autism spectrum disorder

Giorgia Bussu

DONDERS SERIES

Focusing on the individual infant: classification and heterogeneity of autism spectrum disorder

Giorgia Bussu

COLOPHON

The research leading to this thesis was carried out at the Donders Institute for Brain, Cognition and Behaviour at the Department of Cognitive Neuroscience of the Radboud University Medical Centre (Nijmegen, the Netherlands); at Karakter Child- and Adolescent Psychiatry University Centre (Nijmegen, the Netherlands); at the Centre for Brain and Cognitive Development, Birkbeck College, the University of London (London, United Kingdom), and at King's College London at the Psychology Department, Institute of Psychiatry (London, United Kingdom); at King's College London at the Department of Child & Adolescent Psychiatry and MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience (London, United Kingdom); and at Mentis Cura (Reykjavík, Iceland).

The research leading to this thesis was primarily supported by the Marie Sklodowska Curie Actions of the European Community's Horizon 2020 Program under grant agreement n°642996 (Brainview).

The work presented in this thesis was further supported by the UK Medical Research Council (G0701484 & MR/K021389/1); the BASIS funding consortium led by Autistica (www.basisnetwork.org), and EU-AIMS (the Innovative Medicines Initiative joint undertaking grant agreement no. 115300, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution); a Tuberous Sclerosis Association fellowship and the NIHR Maudsley Biomedical Research Centre (BRC) at the Institute of Psychiatry, Psychology & Neuroscience and South London and Maudsley NHS Foundation Trust.

ISBN: 978-94-6284-183-3
Cover: Ilse Modder, www.ilsemodder.nl
Lay-out: Ilse Modder, www.ilsemodder.nl
Printed by: Gildeprint Eschede, www.gildeprint.nl

MIX
Paper from responsible sources
FSC® C012854

Focusing on the individual infant: classification and heterogeneity of autism spectrum disorder

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 20 juni 2019
om 10:30 uur precies

door

Giorgia Bussu
geboren op 2 oktober 1990
te Civitavecchia (Italië)

Promotoren: Prof. dr. Jan K. Buitelaar
Prof. dr. Christian F. Beckmann

Copromotor: Dr. Emily J. H. Jones (Birkbeck College, Verenigd Koninkrijk)

Manuscriptcommissie: Prof. dr. A.H.N. Cillessen
Prof. dr. M.A.J. van Gerven
Prof. dr. C. Kemner (Universiteit Utrecht)

"All that is gold does not glitter,
Not all those who wander are lost;
The old that is strong does not wither,
Deep roots are not reached by the frost."

J.R.R. Tolkien

TABLE OF CONTENTS

Chapter 1	General introduction	11
Part I:	**Early prediction of ASD outcome**	**28**
Chapter 2	Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis	31
Chapter 3	Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants	77
Chapter 4	Widespread atypical neural responses to faces at 8 months predicts autism spectrum disorder at 3 years	103
Part II:	**Parsing heterogeneity in autism spectrum disorder**	**140**
Chapter 5	Latent trajectories of adaptive behaviour in infants at high and low familial risk for autism spectrum disorder	143
Chapter 6	Uncovering neurodevelopmental paths to autism spectrum disorder through an integrated analysis of developmental measures and neural sensitivity to faces	167
Chapter 7	Summary and general discussion	195
Appendix	References	215
	Nederlandse samenvatting	237
	Acknowledgements	243
	About the Author	247
	List of publications	249
	Donders Graduate School for Cognitive Neuroscience series	251

1

General introduction

OVERVIEW

Social communication and interaction are essential to the human condition; some individuals present, however, with a profound impairment in these capacities. The most striking example thereof is Autism Spectrum Disorder (ASD), considered to be among the most severe neurodevelopmental disorders in terms of prevalence, morbidity and impact on society. This thesis focuses on the early signs of and precursors to ASD in the first years of life, and the cognitive and brain mechanisms involved. Figure 1 illustrates an overview of the thesis structure and the experimental chapters. In this introductory chapter, I will first provide background information on ASD; then I will describe the prospective high-risk design, as a powerful research approach to investigate early signs for ASD in infant siblings, and I will report findings in different behavioural (early symptoms, cognitive development, adaptive functioning and temperament) and biological (face processing) domains; finally, I will describe current research on early prediction of ASD, focusing on its challenges (heterogeneity) and methods (supervised classification and unsupervised learning). At the end, I will present the aims of the thesis and the outline of the different chapters.

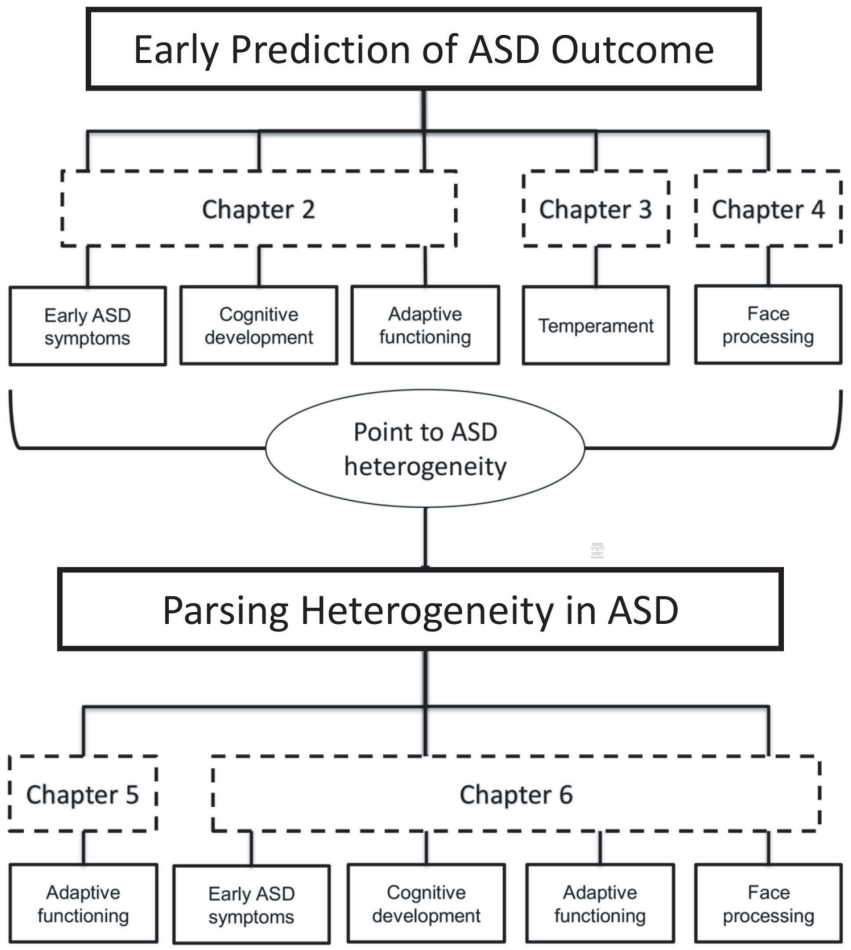


FIGURE 1: Overview of the experimental chapters.

AUTISM SPECTRUM DISORDER

The first description of autism as a syndrome can be traced to 1943, when Leo Kanner[5] and Hans Asperger[6] used the term autism to define behaviours that they separately observed in children. Kanner syndrome was characterized by symptoms like social withdrawal, desire for sameness, communication/language impairment, stereotyped motor behaviours, and intellectual disability with onset from the first year of life. Asperger instead described autistic tendencies in children which differed from what Kanner described by the expression of exceptional isolated talents and conserved linguistic abilities.

Currently, Autism Spectrum Disorder (ASD) is defined as a set of heterogeneous developmental disorders characterised by difficulties in the social-communication domains, restricted repetitive patterns of behaviours and interests, and sensory abnormalities[8]. A diagnosis of ASD according to the DSM-5 criteria requires that symptoms cause clinically significant impairments in social, occupational, or other important areas of functioning. In its heterogeneous manifestations, ASD is viewed as a multifactorial disorder that is due to multiple genetic and environmental factors, and their interaction. Hundreds of genetic variants[10], both as common and rare variants, have been identified that contribute to the development of ASD by impacting the regulation of fundamental processes of early brain development such as cortical organization, synapse structure and function, neural structural and/or functional connectivity, and/or the excitation/inhibition balance[11, 12]. Furthermore, ASD has been linked to monogenetic diseases[13, 14] such as fragile X syndrome, tuberous sclerosis complex and neurofibromatosis type 1 (NF1). On the other hand, we can list gestational diabetes, prenatal drug exposure, preterm birth, congenital infection and neonatal hypoxia conditions among environmental and medical risk factors[15].

ASD is one of the most common neurodevelopmental disorders, with a population prevalence estimated to be between 1% and 1.5%[16, 17]. While some children diagnosed with ASD are able to function independently later in adulthood, most of them are unable to achieve a positive outcome[18, 19] and require life-long care and support. Lifetime costs of services and lost productivity have a huge emotional and economic impact on patients and their families[20]. The best prognosis for ASD currently lies in early targeted intervention aimed to improve later outcome by modifying emergent atypical developmental trajectories[21, 22]. In fact, earlier access to intervention can have long term benefits on functional outcome for children with ASD but also allow costs reduction due to a reduced need of services later in life[23-25]. Thus, considerable research efforts have been focused on the investigation of early manifestations and early diagnosis of ASD.

EARLY SIGNS FOR ASD: THE HIGH-RISK DESIGN

By investigating early signs of ASD, a series of key questions emerges, such as: how do different early developmental trajectories converge into the developmental pathway leading to an ASD diagnosis? Which tools in developmental neuroscience would allow detection of ASD even before a diagnosis is made?

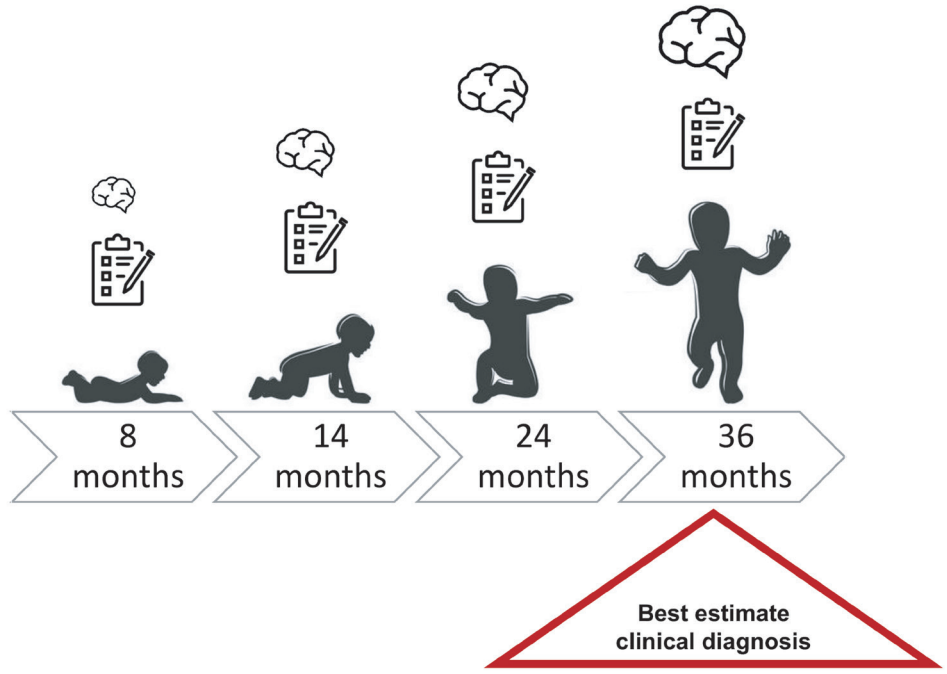


FIGURE 2: prospective design. This figure illustrates the scheme of a prospective design as used for longitudinal studies on infant siblings. Infants are followed up longitudinally over several visits until a best estimate diagnosis is assigned (e.g. at 36 months of age). During the visits, infants' development is assessed through different behavioural and brain measures.

Although ASD symptoms typically emerge early in life, the mean age at diagnosis is still between 4 and 5 years[26]. To provide insight into the emergence of ASD and potentially allow early recognition and diagnosis of ASD, research has been focused on prospective longitudinal studies of infants at high-risk (HR) for ASD based on having an older affected sibling. In fact, high-risk infants have about a 20% risk of developing ASD, significantly higher than the population prevalence[27-29]. In this high-risk design, younger siblings of children

with ASD are followed from baby age to at least 3 years of age, when they receive a clinical evaluation and eventually a best estimate clinical diagnosis of ASD (Figure 2). On the other hand, low-risk controls (LR), namely infant who do not have an older sibling with ASD, are usually also included as a comparison group in these studies. Early measures from high-risk sibling can be compared to low-risk controls to evaluate family-genetic contributions to the disorder, informing on the *autism trait*, while measures from high-risk siblings who later received an ASD diagnosis and their typically developing peers can be compared to assess specific effects leading to ASD as opposed to non-ASD outcome, informing on the *autism state*. This approach allows to understand the behavioural, cognitive and neural mechanisms that precede the clinical onset of ASD and to investigate early manifestations of the condition in a way that is less affected by atypical interactions with the social and physical environment[30]. In this framework, results from previous studies have provided evidence for different early behavioural and biological markers for ASD[29-33].

BEHAVIOURAL SIGNS OF ASD

Prospective longitudinal studies focused on behavioural observations suggest the emergence of an ASD prodrome in the first year of life[34], starting with impairments in the sensory and motor domains at or even before 6 months and moving to the social-communication domain around 12 months[35, 36]. Objectively measured behavioural signs emerging before 12 months include a fall in fixation to the eye region between 2 and 6 months[37], reduced gaze fixation to people at 6 months[38], and vocal atypicalities[39]. However, atypical behaviours in the first year of life are often subtle, transient, outside the core domains characteristic of ASD, and often more sensitive to risk status than ASD outcome[35, 40, 41]. Overt, behavioural signs of ASD begin to manifest in the second year of life, as shown by atypical eye contact, visual tracking, disengagement of visual attention, orienting to name and reduced shared positive affect[30, 31, 34, 42-44]. Nevertheless, it is possible that difficulties in identifying behavioural markers for ASD in the first year of life are caused by limitations of the repertoire of measures available[32].

Among the behavioural measures available in the first years of life, this thesis focuses on early ASD symptoms, cognitive development, adaptive functioning and temperament.

Early ASD symptoms

Putative behavioural signs of ASD can be detected before 18 months by a semi-structured behavioural assessment, the Autism Observation Scale for Infants (AOSI, [3]). After 18 months, the Autism Diagnostic Observation Schedule (ADOS, [4, 45]) can be used as a standardized diagnostic instrument for the assessment of communication, social interaction, play and restricted and repetitive behaviours.

The observation of early clinical symptoms of ASD using the AOSI could differentiate high-risk siblings developing ASD at 24 months from their non-ASD peers at 12 but not at 6 months[46]. Subsequent studies have shown that the AOSI can differentiate high-risk siblings developing ASD from low-risk controls in visual tracking and social referencing at 7 months; however, differences in total scores missed significance[47]. Nevertheless, there was a gradient of AOSI items-level (visual tracking) and total scores at 7 months, and items-level scores (orientation to name) at 14 months in line with the concept of the Broad Autism Phenotype (BAP, [48]); going from the lowest total scores in low-risk controls, higher scores in high-risk siblings not developing ASD, and the highest scores in high-risk siblings developing ASD.

Cognitive development

Verbal and non-verbal cognitive development can be measured by the Mullen Scales of Early Learning (MSEL)[1], a standardized developmental measure assessing cognitive functioning between birth and 68 months in 5 main scales: gross motor skills, visual reception skills, fine motor skills, and receptive and expressive language skills.

Recent reviews of high-risk sibling studies report clear evidence for divergence of siblings developing ASD from typically developing siblings in cognitive and motor skills from 12 months onwards[30, 31]. A prospective study on infants at high and low familial risk for ASD explored trajectories of cognitive development measured by the MSEL and differentiated the infants in 3 groups: infants meeting criteria for ASD at 24 months, infants showing language delay (LD) and unaffected infants[49]. Groups were not significantly different at 6 months, but the ASD group performed significantly worse than the unaffected group on all scales except for visual reception by 14 months and significantly worse than the LD group in gross motor, fine motor and receptive language scales by 24 months. The ASD group did develop slower than the other groups, with the largest slowing in motor domains, and deficits in motor skills by 14 months were associated with later diagnosis of ASD[49]. Such an association has been shown in other studies[50, 51], even already by the age of 7 months[52, 53]. Later studies have replicated and extended these findings on cognitive profiles[53-56], reporting poorer performance on verbal (particularly receptive language) relative to nonverbal skills in high-risk infants developing ASD[56]. This provides supporting evidence to a vulnerable period of progressive divergence of high-risk siblings who later meet criteria for ASD from typical development between 14 and 24 months affecting language, social and motor development.

Adaptive functioning

Adaptive behavior assesses the ability of an individual to function independently in

everyday situations[2]. It can be measured by the Vineland Adaptive Behavior Scales (VABS-II)[2], which is a standardized clinical instrument administered as a semi-structured parent-report questionnaire before 24 months and a parent interview afterwards. Measures are reported in 4 different domains: Communication, Daily Living Skills, Socialization and Motor Abilities.

Adaptive behavior as an outcome measure has been under-investigated, being mostly used as a contributor to the diagnostic decision[30]. However, previous studies have shown lower levels of adaptive behavior in HR siblings at 20[57] and 36 months[58] compared to LR controls[35, 55]. A recent study from the BASIS Team[59] examined developmental trajectories of cognitive and adaptive behaviour abilities between 7 months and 7 years of age in infants at high and low familial risk for ASD. This study compared the low-risk infant group with high-risk groups based on clinical outcome at 36 months (ASD and non-ASD siblings), and showed that HR siblings who go on to develop ASD have increasing difficulties in adaptive behaviour over time compared to LR controls, while it's not the case for HR siblings who do not develop ASD.

Temperament

Temperament can be defined as individual differences in activity, affectivity, attention and self-regulation that are adapted throughout development by complex interactions between genetic, biological and environmental factors[60]. These differences are already measurable at an early age under three main domains: surgency/approach, referring to engagement with the environment, positive emotions and activity level; negative affect/withdrawal, including negative emotions such as anger, sadness and fear; and effortful control, referring to regulation of attention, emotions and behaviors[61].

Several studies have focused on the association between early temperament and the development of ASD. Low levels of surgency have been associated with ASD from 24 months onwards[46, 62, 63], while in the first year of life, HR siblings developing ASD have been shown to have higher levels of surgency than their non-ASD peers[64, 65]. On the other hand, more difficult effortful control has been shown in HR siblings developing ASD from 12 months onwards[63, 65, 66]. Findings on negative affect are mixed instead, with studies showing an association with ASD from 6 months onwards[46, 65, 67, 68] but also no differences in negative affect between HR siblings developing ASD and LR controls in early childhood[62, 64].

BIOLOGICAL SIGNS OF ASD

Differences in brain development appear to precede changes in behavior; thus, the observation of early brain development holds the potential to detect signs of ASD before behavioural signs and clinical symptoms emerge[32]. Prospective imaging studies on high-risk siblings have identified atypical patterns of brain growth, structural and functional connectivity, and atypical brain responses to different stimuli in infants developing ASD ([32, 69, 70] for reviews). Among these early biological signs of ASD, this thesis focuses on face processing as a precursor of social skills development.

Face processing

Electroencephalography (EEG) and event-related potentials (ERPs) provide a useful tool to examine the neural correlates of face recognition in infancy[71]. Both EEG and ERP measure electrical activity of the brain recorded from scalp electrodes, and more specifically, ERPs reflect changes in the electrical activity of the brain time-locked to a specific event (e.g. an external stimulus or a mental state). ERPs do not require an overt behavioural response, thus allowing to study the relation between brain and behavior beginning at birth[72]. In comparison to other brain imaging tools like MRI, EEG and ERP provide high temporal resolution (in the order of milliseconds), but low spatial resolution[73].

The ERP waveform in response to faces in infancy typically presents the characteristic P1, N290, and P400 components, considered to be infant precursors of the face-sensitive N170 in older children and adults[74-76]. These components are known to be modulated by face perception, and more specifically direction of eye-gaze, as early as 4 months of age[77]. The P1 is modulated by selective attention and reflects lower level, sensory processing. Higher P1 amplitude in response to faces early in development has been associated with greater social interest and approach later on, suggesting an association between early sensory sensitivity and social development[78]. On the other hand, later components, such as the P400, reflect higher level cognitive processing related to retrieval of semantic information related to faces[76, 79].

Infants developing ASD demonstrate emerging atypicalities in social-communicative behaviour from the first year of life, with a declining interest in human faces[37, 80, 81] by 6 to 12 months of age. These behavioural findings are accompanied by atypical neural responses to faces as measured by ERPs[82-84]. Previous studies have shown a delayed profile of N290 modulation by gaze in children with ASD, indicative of delay in development of mechanisms underlying gaze processing[85]. Further, preschool children with ASD show a lack of P400 amplitude differentiation between familiar and unfamiliar faces but not objects[82, 84] and delays in the speed of the N290 response to faces that are associated with their social developmental level[84]. Infant sibling studies have indicated that atypicalities in these components could emerge in early development. HR siblings developing ASD have shown

a more rapid P400 peak response to faces versus objects at 6 months compared to their typically developing peers[86], and reduced differentiation in P400 amplitude responding to faces that shift gaze towards versus away from the viewer compared to non-ASD siblings between 6 and 9 months[87]. Similarly, atypical neural responses to social stimuli in high-risk siblings were associated to difficulties in socialisation in toddlerhood[88].

EARLY PREDICTION OF ASD

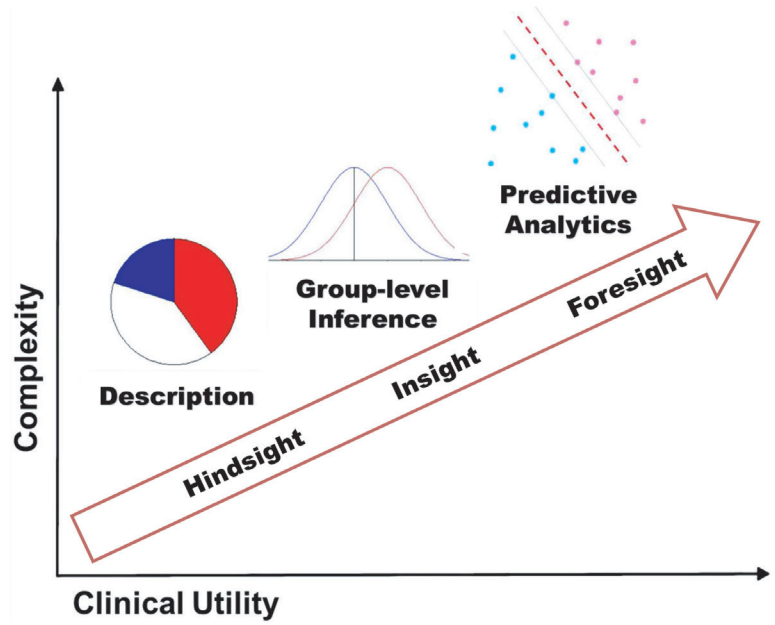


FIGURE 3: predictive analytics. Early detection of ASD shifts the focus from the description of patients (hindsight) and the investigation of statistical group differences or associations (insight) toward models capable of predicting future characteristics for individual patients (foresight). This figure was adapted from Hahn et al.[89].

Research on intervention with infants at high-risk for ASD suggests that behavioural intervention might be more effective when delivered within the time window of age 7 to 15 months, when the core symptoms of ASD have not emerged yet[90], rather than later[91, 92]. However, the sustained delivery of behavioural intervention to all infants at risk for ASD based only on traits would be too expensive, and the risk/benefit ratio may be less favourable for infants who would have developed typically anyway. Thus, while intervention can be performed based on *traits* only if it does not bring disadvantages to false-positives and it is low-cost, individual prediction of ASD in the first year of life might be crucial to identify the infants who need intervention and enable early targeted intervention. This is why research on early detection of signs of ASD has started to focus on the translation of group differences to individual prediction of ASD outcome at early age (Figure 3). In fact, group differences do not tell us anything about the individual infant as they can be

significant despite an overlap between groups in individual variation. Thus, moving from group-level to individual-level analyses is essential to understand individual variability and allow prediction at the level of the individual infant[89, 93].

HETEROGENEITY IN ASD

ASD is a clinically heterogeneous disorder[94]. This intrinsic heterogeneity is evident on different levels of analysis (e.g. onset, aetiology, symptom profiles, severity) and present both between individuals under the clinical label of ASD and within individuals across development[95-97]. This points to multiple underlying processes acting together across multiple functional domains and leading to the disorder, rather than a unitary biological process[30, 94, 98, 99]. Thus, to understand the neurodevelopmental mechanisms leading to ASD it is essential to decompose this heterogeneity[100], which requires in its turn multiple levels of analysis within and between individuals across development[93, 101]. In this perspective, the integration of different types of data is essential to combine complementary information across domains and capture ASD multiplicity, while longitudinal measures are fundamental to investigate developmental changes over time within and between individuals.

Few studies so far have conducted analyses that combined measures from different domains. Estes and colleagues[35] integrated MSEL, VABS and AOSI data to investigate early developmental characteristics of children at high and low risk for autism in relation to ASD outcome at 24 months. Results showed a pattern of symptoms starting in the sensorimotor domain at 6 months and moving to abnormalities in the social and communication domain after 12 months. As for prediction of ASD outcome, previous studies have integrated data from different modalities, like structural, diffusion and spectroscopy magnetic resonance data[102] or functional neuroimaging and behavioural (ADOS) data[103], showing an increase in predictive power for ASD.

MACHINE LEARNING FOR CLASSIFICATION OF ASD OUTCOME

Limits to early detection of ASD at an individual level come from the high clinical, biological and etiological heterogeneity of the disorder[94]. The investigation of different types of data is essential to more comprehensively capture the different aspects of the disorder, and machine learning holds the potential to provide a robust multivariate model for prediction of later clinical outcome combining complementary information from different sources in an efficient way, and objectively identify disorder-specific features in data allowing for automatic case-control classification[104, 105]. In addition to clinical utility as potential screening tools or to provide complementary information for difficult diagnoses[105], machine learning tools can be used in psychiatry to develop meaningful theories for the disorder of interest[106]. In fact, a robust model for the relevant clinical pathology should be able to

differentiate cases from controls at an individual level.

Classification is the process of taking some input measures (features) for a series of cases and integrate them into the assignation of a binary label (class) to each case. Thus, it can potentially be used to differentiate infants who go on to receive an ASD diagnosis from those who don't. To test generalizability of a classifier, cross-validation is performed by splitting the available dataset into a 'training set' and a 'test set'. Model determination and parameter estimation is performed on the 'training set', for which labels are known in supervised learning. Next, predictive performance of the classifier is evaluated on the 'test set', for which labels are unknown to the classifier. Finally, to determine clinical utility of the classifier, it is essential to test performance reliability on an independent dataset[105]. Generalizability of the classifier can be hampered by overfitting, which is a condition in which the model perfectly fit the training data, including noise, leading to fitting error close to zero but much higher in independent data[107]. To reduce overfitting, it is important to perform cross-validation and balance model complexity with sample size.

The application of these methods has already shown promise for classification of children with ASD[108-117]. Previous studies have used a machine learning approach on behavioural measures to predict ASD outcome and an individual level[113, 114]. In particular, Chawarska and colleagues[114] used measures from the ADOS to predict ASD outcome at 36 months in a cohort of high-risk siblings at 18 months. A nonparametric decision-tree learning algorithm[118] was applied to standardized behavioural ratings to identify the individual items of the ADOS at 18 months that best differentiated high-risk siblings developing ASD from typically developing siblings or siblings with other developmental disorders. A combination of 6 behavioural features (i.e. unusual eye contact, unusually repetitive interests or stereotyped behaviours, intonation of vocalizations or verbalizations, giving, imagination or creativity, gestures) allowed the identification of ASD cases with 83% accuracy, while social-communicative and other autism-specific features alone did not provide a good predictive value for ASD. This suggests that the *interaction between (or combination of)* individual behaviours must be considered to enhance their predictive value for an early identification of later ASD outcome. Furthermore, this approach shed light on different sibling subgroups at 18 months, characterized by distinct developmental pathways with respect to ASD symptoms severity and verbal and non-verbal skills, identifiable via different combinations of features: the first subgroup was characterized by limited nonverbal communication and marked symptoms between 18 and 36 months; the second by repetitive behaviours and mild symptoms which intensify by 36 months; the third by moderate symptoms with poor eye contact and a limited ability of spontaneous pretend play.

While classification of ASD from behavioural measures before 12 months has not been reported, previous studies have shown that brain data are more successful at predicting individual ASD outcome in the first year of life[111, 112]. Emerson and colleagues[111] used a support vector machine algorithm[119] on patterns of infant brain functional connectivity during natural sleep at 6 months to predict ASD outcome at 24 months in HR siblings, reaching 97% accuracy. Similarly, Hazlett and colleagues[112] used a neural network[120] on structural magnetic resonance imaging (MRI) data from 6 and 12 months of age to predict ASD outcome at 24 months in a HR sample, reaching a classification accuracy of approximately 94%. However, results need to be replicated in larger, independent samples to test for generalizability and potential clinical utility, while prediction of a more stable ASD diagnosis at 36 months must be also tested.

UNSUPERVISED LEARNING

However, a limitation to the understanding of ASD heterogeneity might also come from the traditional case-control comparison approach. In fact, although valuable to identify potential early risk markers for ASD, the classic case-control comparison approach is based on predefined clinical labels to partition the sample in clinical groups. This supervised approach does not take into account the heterogeneity of psychiatric disorders like ASD, which often share symptoms and might be better understood as a continuous spectrum rather than separate entities[121]. Furthermore, in the context of prospective designs, the case-control comparison approach assumes that the clinical label defined at the time of diagnosis has meaning earlier in development. This points to the necessity to explore different methods that might be more prone to capture unknown structure in data and help refining categorical outcomes.

Unsupervised learning is a hypothesis-free approach to unlabelled data for the inference of structure, thus allowing to identify intrinsic patterns in data without any a priori knowledge on the sample subgrouping. In this thesis, I employed growth mixture modelling for stratification based on developmental trajectories (Chapter 5), as opposed to traditional cross-sectional clustering[122], and linked independent component analysis (ICA; Chapter 6) to identify the underlying processes associated with clinical outcome based on the extraction of intrinsic patterns in multivariate data.

Growth mixture modelling

Growth mixture modelling (GMM) is a person-centred approach to longitudinal modelling which allows to classify individuals into separate groups based on individual trajectories of responses, with individuals within a group being more similar than individuals between groups[123]. In conventional growth modelling, random effects on intercept and slope at an

individual level allow to capture heterogeneity of growth curves[124]. However, a subset of individuals might show significantly different growth trajectories than the average estimate. GMM does not assume common growth parameters across the entire sample; rather, it allows to identify different growth parameters across unobserved subgroups in the sample. The model consists of a simultaneous estimation of a mixture model growth for class-varying random coefficient means and individual classification for class membership. In comparison, latent class growth analysis (LCGA) is a related method used to model unobserved heterogeneity in longitudinal data[123]; however, it assumes that all individuals in a subgroup have homogeneous trajectories while GMM allows for within-class variation[125]. Thus, especially with complex data, GMM provides a more realistic representation of heterogeneity in growth trajectories, allowing to capture the complexity of developmental variation across individuals and to better characterize the developmental process itself.

Linked ICA

Linked ICA allows to simultaneously model and discover independent sources of signal across multiple modalities by seeking non-Gaussianity in data[126-128]. Although mainly used with neuroimaging data[129-132], this method can be applied to integrate data from different modalities collected on the same participants in large cohorts of infant siblings. In fact, modalities can be linked by the participant loading matrix, shared across modalities, whereby each identified component will be defined by a single participant loading and a map of scores per modality. This allows to potentially identify underlying biological processes leading to ASD development that might have an effect on multiple measures from different modalities. Furthermore, it allows to simultaneously model data from the same participants collected at different time points and extract components linked across time, informing on the developmental effect that those underlying mechanisms might have longitudinally.

THESIS OUTLINE

In this thesis, I investigate predictive power of different measures collected in the first three years of life for prediction of ASD diagnosis at the level of the individual infant, and explore data analysis strategies for the investigation of ASD heterogeneity. My main hypotheses are that the integration of data from different modalities and different time points could: (1) improve predictive power for detection of ASD in the first year of life; (2) allow stratification of ASD heterogeneity. Here, I focused on the individual rather than clinical groups. All the analyses reported in these studies have been performed on the same sample (see Box 1.1). The first part of the thesis focuses on prediction of ASD outcome. Chapter 2 illustrates the analysis of longitudinal measures of developmental level, adaptive functioning and ASD symptoms between 8 and 36 months, showing developmental trajectories by outcome groups and using different combination of scores at 8 and 14 months to predict individual ASD and more general 'atypical' outcome at 36 months. Similarly, Chapter 3 illustrates the analysis of temperamental factors between 8 and 36 months, integrating group-based and individual-level analyses to investigate predictive power of these scores for ASD and 'atypical' outcome. Then, in Chapter 4 I move from behavioural and developmental data to brain data, investigating predictive power for ASD coming from neurophysiological responses to social and non-social stimuli in high-risk siblings at 8 months. Although prediction had high accuracy (approximately 80%) using neural sensitivity to faces as opposed to visual noise at 8 months, these initial studies pointed to the high inter-individual heterogeneity in ASD and different mechanisms underlying ASD development.

The second part of the thesis focuses on the investigation of ASD heterogeneity through unsupervised, data-driven analyses. Chapter 5 illustrates classes of developmental trajectories of adaptive functioning between 8 and 36 months in infants at high and low familial risk of autism identified through growth mixture modelling. Next, Chapter 6 illustrates the uncovering of neurodevelopmental associated to clinical outcome at 36 months using measures of neural sensitivity to eye gaze, developmental level, adaptive functioning and early symptoms of ASD in high and low risk infants at 8 months.

In conclusion, this thesis shows the first steps in multi-modal multi-domain data integration for classification of ASD at an individual level in the first years of life and prospective investigation of its heterogeneity. Results indicate that brain data, as opposed to behavioural and developmental data, retain the highest value for early detection of ASD. However, all these studies also point to the complexity of ASD in its manifestations and the necessity to decompose its heterogeneity to improve the understanding of its underlying mechanisms and allow a more accurate prediction of outcome at an early age.

BOX 1.1: BASIS sample.

Data presented in this thesis were collected from infants recruited across two phases (Phase 1 and Phase 2) of the British Autism Study of Infant Siblings (BASIS, http://www.basisnetwork.org), a large longitudinal study on infants at high (HR) and low familial risk (LR) for ASD, for a total of 247 infant participants. HR siblings were infants with at least an older sibling with an ASD diagnosis, while LR controls were full-term infants recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. A cohort of 54 HR siblings and 50 LR controls participated in Phase 1, and an independent cohort of 116 HR siblings and 27 LR controls participated to Phase 2.

Multiple measures were collected during 4 visits at approximately 8, 14, 24 and 36 months of age. The Mullen Scales of Early Learning (MSEL; [1]) and the Vineland Adaptive Behavior Scale (VABS; [2]) were administered at all visits to assess infant's developmental level and adaptive functioning. The Autism Observation Scale for Infants (AOSI, [3]) was administered at 8 and 14 months to obtain information about ASD symptomatology, while the Autism Diagnostic Observation Schedule (ADOS; [4]) was administered at 24 and 36 months, and the Autism Diagnostic Interview Revised (ADI-R; [7]) at 36 months. At 8 months, electrophysiological responses to face/gaze were also measured.

At 36 months, infants (n=239) were seen for a clinical evaluation. Expert clinical researchers reviewed all available information at 24 months and 36 months and assigned clinical consensus best estimate diagnosis of ASD according to ICD-10 [Phase 1][9] or the then published DSM-5 criteria [Phase 2][8]. The best estimate diagnoses for the cohorts from the two different phases were reviewed for differences in categorisation between samples and considered to be similar.

Overall, no formal clinical diagnoses were assigned to the LR group, which was only based on risk sampling assignation. However, none of the LR infants raised concerns for atypical development. Among HR infants who did not meet criteria for ASD, a subgroup of siblings was classified as 'atypical' based on having: ADOS and/or ADI-R above ASD threshold, and/or MSEL more than 1.5 standard deviations below average on receptive language and/or expressive language and/or early learning composite.

All procedures were in agreement with ethical approval granted by the London Central NREC (approval code 06/MRE02/73), and one or both parents gave informed consent to participate in the study.

PART I

Early prediction of ASD outcome

2

Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis

Published as: "Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis", J Autism Dev Disord, 2018.*

Bussu, G.,
Jones, E. J. H.,
Charman, T.,
Johnson, M. H.,
Buitelaar, J. K. & The BASIS Team

*Adapted from the original form for this thesis.

ABSTRACT

We integrated multiple behavioural and developmental measures from multiple time-points using machine learning to improve early prediction of individual Autism Spectrum Disorder (ASD) outcome. We examined Mullen Scales of Early Learning, Vineland Adaptive Behavior Scales, and early ASD symptoms between 8 and 36 months in high-risk siblings (HR; n=161) and low-risk controls (LR; n=71). Longitudinally, LR and *HR-Typical* showed higher developmental level and functioning, and fewer ASD symptoms than *HR-Atypical* and *HR-ASD*. At 8 months, machine learning classified *HR-ASD* at chance level, and broader atypical development with 69.2% Area Under the Curve (AUC). At 14 months, ASD and broader atypical development were classified with approximately 71% AUC. Thus, prediction of ASD was only possible with moderate accuracy at 14 months.

INTRODUCTION

Although symptoms of Autism Spectrum Disorders (ASD) typically emerge early in life, a reliable diagnosis is usually not achieved before age 3 or later [26]. Evidence suggests that the best prognosis for ASD currently lies in early targeted intervention aimed to improve later outcome by modifying emergent atypical developmental trajectories [21, 22]. A recent follow-up study on the effects of parent-mediated social communication intervention in infants at high familial risk of ASD between 9 and 14 months shows a treatment effect on symptom severity extended 24 months after intervention end [133]. However, the sustained delivery of behavioural intervention to all infants at risk for ASD based only on traits would be too expensive, and the risk/benefit ratio may be less favourable for infants who would have developed typically anyway. Thus, individual prediction of later development of ASD as soon as early signs emerge could help to better target early intervention strategies.

Limits to detection of ASD before 24 months come from the high heterogeneity of the disorder and the relatively late emergence of the core characteristics of ASD. Heterogeneity in onset, aetiology, phenotype, neurobiology, and developmental trajectory points to multiple underlying processes acting together and leading to the disorder rather than a unitary biological process [94, 99, 134]. Therefore, the investigation of different types of data is essential to capture the different aspects of the disorder. Machine learning holds the potential to provide a robust algorithm for prediction of later clinical outcome combining complementary information from different sources in an efficient way, and allowing the identification of the most predictive combination of measures. The application of these methods has already shown promise for classification of children with ASD [108-110]. In the present study we apply machine learning algorithms to predict clinical outcome at 36 months from different combinations of behavioural and developmental measures at 8 and 14 months. Despite a general consensus on the added value of data integration for prediction of ASD, the method has not previously been applied to behavioural measures and standard developmental assessments.

Research in the early recognition and diagnosis of ASD has been focused on prospective longitudinal studies of infants at high-risk for autism because they have an older sibling with ASD. High-risk infants (HR) have about a 20% risk of developing ASD, significantly higher than the population prevalence of 1.5% [16, 27-29], and thus the high-risk design allows us to study the early manifestations of the condition and understand the behavioural, cognitive and neural mechanisms that precede the clinical onset of ASD [134]. Yet, these studies have mainly focused on average differences between infants who later develop ASD and their typically developing peers, measuring group differences by means of p-values. Convergent

evidence supports the emergence of overt behavioural markers for ASD by the end of the second year of life, such as atypical eye contact, visual tracking, disengagement of visual attention and orienting to name [31, 34, 42-44, 134, 135]. Objectively measured behavioural signs for ASD emerging before 12 months include a fall in fixation to the eye region between 2 and 6 months [37], reduced gaze fixation to people at 6 months [38], and vocal atypicalities [39]. But early markers for ASD are not limited to social domains. By 14 months, high-risk siblings developing ASD performed significantly worse than unaffected siblings on all scales of the Mullen Scales of Early Learning (MSEL) except for Visual Reception [49], and impairments in verbal skills (particularly receptive language) [56], and motor skills were associated with later diagnosis of ASD [49-51, 54], even already by the age of 7 months [52, 53]. Few studies so far have conducted analyses that combined measures from different domains. Estes and colleagues [35] investigated trajectories of developmental abilities, as measured by the MSEL; adaptive functioning, as measured by the Vineland Adaptive Behavior Scales (VABS); and early ASD symptoms, as measured by the Autism Observational Scale for Infants (AOSI), in infants at high and low risk for autism in relation to ASD at 24 months, showing a pattern of symptoms starting in the sensorimotor domain at 6 months and moving to the social-communication domain after 12 months. Thus, prediction of autism may require a multi-measure approach.

Although previous findings on group differences between infants who later develop ASD and their typically developing peers are valuable in terms of finding relevant biomarkers for the disorder, there is often substantial overlap between groups in individual variation, making prediction for individual infants difficult. The aim of individual prediction of outcome is to automatically classify each individual into one group (e.g. ASD vs. non-ASD outcome), and performance is usually measured by accuracy or Area Under the Curve (AUC). The AUC is a measure of predictive accuracy computed as the area under the Receiver Operating Characteristic (ROC) curve, which is a plot of true positive rate vs. false positive rate for the model under evaluation [136]. Prediction at chance level results in 50% AUC, while prediction has moderate accuracy for AUC above 70%. Few studies have used behavioural measures to predict individual outcome of ASD, individual prediction being more focused on neuroimaging data [102, 104, 111, 137]. Macari and colleagues [113] employed a decision-tree nonparametric learning algorithm to classify typical versus 'atypical' high-risk infants using as features measures from the Autism Diagnostic Observation Schedule (ADOS) at 12 months. 'Atypicality' included but was not limited to ASD, and was based on clinical evaluation at 24 months. Despite promising results, the study was considered preliminary due to a small sample size (n=84) and the lack of a confirmatory diagnosis at 36 months. Chawarska and colleagues [114] used the same methods to predict ASD outcome at 36 months in a cohort of high-risk siblings at 18 months. The aim was to identify the

individual items of the ADOS-G at 18 months that best differentiated high-risk siblings who were going to develop ASD from typically developing siblings or siblings with other developmental disorders. The combination of 6 behavioural features (i.e. repetitive behaviours, eye contact, intonation, gestures, giving objects and spontaneous pretend play) allowed the identification of ASD with high accuracy (83%), while poor eye contact or limited gestures alone did not provide good prognostic value for ASD. This suggests that the *interaction between (or combination of)* individual behaviours must be considered to enhance predictive value for an early identification of later ASD outcome. Prior to our study, classification of ASD from behavioural measures before 12 months has not been reported, while it would be crucial to enable early intervention. Furthermore, previous studies only looked at items from the ADOS, but did not investigate whether different measures of developmental skills and functioning can increase predictive power for ASD at an early age.

The aim of the present study was to investigate predictive longitudinal differences from 8 to 36 months between infants at low and high familial risk for autism with different developmental outcomes (typical, ASD, atypical). Further, we investigated whether we could predict ASD or atypical development at 36 months at an individual level within the HR group from data collected at 8 and 14 months. Extending the approach adopted in previous studies, we integrated measures from ASD symptoms, developmental and adaptive functioning, and we compared classifiers based on different combinations of measures to identify which combination is most predictive. We tested the hypothesis that integration of information about symptoms, developmental ability and everyday functioning can improve prediction of ASD compared to prediction from ASD-specific symptoms alone, capturing pervasiveness and addressing the high heterogeneity of ASD. Prediction was also made taking into consideration the dynamics of development by adding the change of scores between 8 and 14 months to cross-sectional measures at 8 months. This allowed us to test our second hypothesis that integration of measures from multiple time-points adds value to prediction of ASD from measures at early age compared to prediction from measures at single time-points.

METHODS

Participants

Data presented in the current paper were collected as part of a large longitudinal study, to which 247 infants participated in one of two phases of longitudinal assessments (104 in Phase 1 and 143 in Phase 2). Data from 232 infants (161 [69.4%] high-risk siblings [HR] and 71 [30.6%] low-risk controls [LR]) were included in this study; 10 infants were excluded because they did not receive an ADOS evaluation and/or a clinical outcome evaluation at 36 months; 5 infants were excluded because they did not attend at least one of the visits. HR infants were at increased familial risk because they had an older biological sibling with ASD, while LR controls had an older full sibling with typical development. The sample was balanced in gender (116 males and 116 females), and 85/161 HR siblings (53%) and 31/71 LR controls (44%) were males. We used imputation through expectation maximization to handle missing data (see *Supplemental Material* for details). Analyses were performed on SPSS (http://www.ibm.com/analytics/us/en/technology/spss).

DEVELOPMENTAL ASSESSMENTS

All infants, irrespective of diagnosis and risk group, were followed longitudinally on four visits from an intake evaluation at 8 months [mean=8.1; standard deviation, SD=1.2] with further assessments at 14 months [mean=14.5; SD=1.3], 24 months [mean=25.4; SD=3.1] and 36 months [mean=38.4; SD=2.3]. At each assessment, infants were evaluated on the MSEL and VABS. Autism symptoms were assessed through the AOSI at 8 and 14 months, while the ADOS was used at 24 and 36 months. The Autism Diagnostic Interview – Revised (ADI-R; [7]), a structured parent interview, was also used to assess autism symptoms at 36 months. Experimenters were aware of infants' risk status, but assessments were blind to clinical outcome. At the time of enrolment, none of the infants had been diagnosed with any developmental condition.

MEASURES

Developmental skills

Verbal and non-verbal cognitive development was measured at each visit by the MSEL [1], a standardized developmental measure used to assess cognitive functioning between birth and 68 months. Scores are obtained in 5 scales and 2 main functional domains: the gross motor scale (GM), and the cognitive scales. The cognitive scales are visual reception (VR), fine motor abilities (FM), receptive (RL) and expressive language (EL). The Mullen Scale provides normative scores for each specific scale (average T-score = 50, standard deviation SD = 10) and a single composite score representing general intelligence (Early Learning Composite, ELC; average standard score = 100, SD = 15). The T-scores from the five MSEL scales were included in this study.

Adaptive functioning

The VABS (VABS-II) [2] is a semi-structured parent-report questionnaire (used at 8 and 14 months) or parent interview (used at 24 and 36 months) completed at each visit to assess infant's adaptive behaviour in everyday settings. The items address personal and social functioning in 4 different domains: Communication (Comm), Daily Living Skills (DL), Socialization (Soc) and Motor Abilities (Mot). An Adaptive Behavior Composite (ABC) provides an overall index of adaptive functioning. The standard scores (mean = 100, SD = 15) from the four domains were included in this study.

Early ASD symptoms.

Early autism symptoms were measured at 8 and 14 months by the AOSI [3], a semi-structured observational assessment designed to detect putative behavioural signs of autism in infants aged between 6 and 18 months. In this study a 19 item version of the AOSI was used [138], and the total score obtained from the sum of codes from the different items as an overall evaluation score was included in the analyses. The ADOS [45] was administered at 24 and 36 month but not included in our analyses. It is a standardized diagnostic instrument for the assessment of communication, social interaction, play and restricted and repetitive behaviours in children older than 18 months.

CLINICAL OUTCOME

Expert clinical researchers reviewed all available information at 24 months (including MSEL, VABS and ADOS) and 36 months (including MSEL, VABS, ADOS and ADI-R) and assigned clinical consensus best estimate diagnosis of ASD according to ICD-10 criteria [9] to HR infants recruited in Phase 1. The same process was followed in Phase 2 and clinical consensus on ASD diagnosis was assigned according to the then published DSM-5 criteria [8]. To check for differences in categorisation between samples, the clinical research lead (TC) reviewed the best estimate diagnoses for the two cohorts together with the team members involved in the diagnostic decision-making, and given the lack of precision in definition in ICD-10 criteria for 'broader ASD' (atypical autism, PDD-unspecified, PDD-other), the broad ASD categorisation being used in both Phases was considered to be similar. Among infants who did not meet criteria for ASD, a subgroup of siblings showed atypical scores and was classified as 'atypical'. Criteria for an atypical outcome were: ADOS and/or ADI-R above ASD threshold, and/or MSEL more than 1.5 standard deviations below average on receptive language and/or expressive language and/or early learning composite. Overall, 32/161 (19.9%) HR infants (24 males) met criteria for ASD at 36 months (*HR-ASD*); 43/161 (26.7%) HR infants (23 males) met criteria for atypical developmental (*HR-*

Atypical); the remaining 86 HR infants did not meet criteria for ASD or any developmental condition (*HR-Typical*). No formal clinical diagnoses were assigned to the LR group, which was only based on risk sampling assignation, but none of them had a community clinical ASD diagnosis at 36 months. In particular, no ADI-R was administered to LR in Phase 1, who did not receive an outcome evaluation. In Phase 2, LR infants were administered the ADOS and ADI-R and received an outcome evaluation at 36 months, but none of them raised any concern for ASD or atypical development.

STATISTICAL ANALYSES: AN OVERVIEW

First, four analysis groups were derived based on combined clinical outcome and risk status: *LR* (n=71), *HR-ASD* (n=32), *HR-Atypical* (n=43), and *HR-Typical* (n=86). A fractional rank based inverse normal transformation was applied to all measures and the transformed data met assumptions of normality, except for MSEL GM and RL scores, and VABS DL scores at 8 months ($p<0.05$), AOSI total score at 14 months ($p<0.001$), and MSEL FM scores at 36 months ($p<0.05$). However, the statistical tests used in this study were very robust and insensitive to violations of normality. Second, to identify differences in developmental trajectories, we compared the four groups with respect to longitudinal profiles of single measures from 8 to 36 months using multilevel mixed modelling. Finally, we performed a classifier analysis on single and multiple measures from single and multiple time points to investigate whether integrated information improved prediction of ASD at pre-diagnostic age.

TRAJECTORY ANALYSIS

We used measures of developmental level and adaptive functioning to characterise longitudinal profiles over 4 visits between 8 and 36 months. The main analysis consisted of linear mixed-effect regression, LMER, to model trajectories of each measure at group level after considering effects at the individual level. AOSI Total Score was excluded from these analyses as it was only available at 2 different time-points. In contrast to a more traditional approach, LMER allows to control for the variance explained by random factors without the necessity to aggregate data [139]. Real age and outcome were included as fixed factors, and gender was included as a covariate, while random effects on intercept and slope were modelled on individual level. We compared linear and quadratic models on age to select the best fit for each variable based on chi-squared tests on the log-likelihood values. Then, we investigated the main effects of *outcome* and *age* (and age^2 for quadratic models), and their interaction effects using Wald tests with Satterthwaite approximation for degrees of freedom. Post-hoc Tukey's tests for multiple comparisons were performed for group comparisons and simple main effects analysis. Finally, we characterised trajectories of estimated values per different outcome groups, and 95% confidence intervals were

computed via bootstrap (n=1000 repetitions). Analyses were implemented using the *lme4* software package on R [140].

CLASSIFIER ANALYSIS

Classification is the process of taking some input measures (features) for a series of cases and assigning a binary label (class) to each case. In supervised learning, the classifier, which is an algorithm that implements classification using a specific set of features, is trained on a set of cases with known labels, and its predictive performance is evaluated on a separate test set with labels unknown to the classifier. In this study, we performed a supervised classifier analysis on infants using as features MSEL, VABS and AOSI scores at 8 and 14 months. The distinction made was between *HR-ASD* vs. *HR-Atypical* + *HR-Typical*. In addition, the classification of *HR-ASD* + *HR-Atypical* vs. *HR-Typical* was performed since the differentiation of the atypical group as a whole from typically developing infants might be useful for identifying HR siblings who would benefit from intervention. Low-risk controls were excluded from the classifier analysis since our main aim was to answer the clinically relevant question of predicting ASD outcome among HR siblings.

The algorithm chosen for classification was a least-squares support vector machine. To validate the classifier against overfitting and allow generalizability, we used 40% holdout cross-validation repeated 10 times. This is a variant of *k-fold* cross-validation in which we choose the percentage of splitting between training and test sets, and the *k* number of repetitions of the learning process independently. Analyses were implemented on Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016) using the Matlab toolbox *LS-SVMlab* (http://www.esat.kuleuven.be/sista/lssvmlab). To maintain correctly evaluated predictive performance, the sample partition into training and test set was made with stratification based on outcome, so that the different sets had similar structure. Furthermore, sampling with replacement was performed on the training set to address class imbalance and avoid a wrong identification of model parameters in favour of the majority class. Model parameters were tuned via an inner cycle of 10-fold cross-validation and the tuning parameters were optimized in a Bayesian framework [141]. Features were z-scored before being entered into the classifier to have similar ranges of scores.

To investigate the predictive power of measures across time, we tested different classifiers using measures from different time points: 8 months; 14 months; and 8 months plus the change factor between 8 and 14 months. Then, to determine the best classifier, we computed the AUC and we evaluated each classifier performance via sensitivity, specificity, accuracy, negative predictive value (NPV: true negative over negative predicted cases) and positive predictive value (PPV: true positive over positive predicted cases). 95% confidence

intervals (CI) for each metric were computed to improve reliability of the obtained estimates using bootstrap with n=1000 repetitions for each cross-validation fold, then averaging over folds. To test whether classification accuracy was significantly better than chance level, we computed the *p-value* of AUC for each classifier through a shuffle test. Labels in the test set were randomly shuffled, and pre-trained classifiers were used for prediction on the test set. This procedure was repeated n=1000 times for each cross-validation fold (n=10000 total repetitions) to estimate the null distribution of AUC and test whether classifiers perform significantly better than random. Then, a nonparametric Friedman test was performed on the AUC of different classifiers at each time point separately to test whether accuracy differed using different measures from the same time point as features. When the Friedman test was significant, we performed post-hoc paired Wilcoxon tests between the classifier with highest AUC and other classifiers. The Bonferroni correction was used to account for biasing effects due to multiple comparisons. The same method was used to test whether the same measures at different time points provided different predictive accuracy. Finally, the paired Wilcoxon test was used to test whether predictive accuracy of the best classifiers at different time points was significantly different within the same classification (*HR-ASD vs. HR-Typical* or *HR-ASD + HR-Atypical vs. HR-Typical*), and whether predictive accuracy of the best classifiers at the same time points was significantly different between the 2 different classifications. This allowed us to assess differences in predictive power for ASD at different time points, and to compare predictive accuracy for ASD vs. broader atypicality. Classifier comparisons were performed on SPSS (http://www.ibm.com/analytics/us/en/technology/spss).

RESULTS

PARTICIPANT CHARACTERISTICS

Global descriptive statistics are summarized in Table 1. There was a significant difference of gender per clinical outcome, with more males receiving an ASD diagnosis at 36 months than females (odd ratio for HR males vs. females developing into ASD OR=3.52 [CI: 1.51 to 8.22], $p<0.005$). Outcome groups did not differ from each other in age at any visit.

TABLE 1. Demographic data for high-risk and low-risk groups by 36-month clinical outcome. This table shows gender (count, *n*), age, and developmental and behavioural measures [*mean (standard deviation, SD)*] by clinical outcome group.

	Overall	High-Risk (n = 161)			Low-Risk(n = 71)
		ASD (n = 32)	Atypical (n = 43)	Typical (n = 86)	
Gender	**N**	**n**	**n**	**n**	**n**
Male	116	24	23	38	31
Female	116	8	20	48	40
Age	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**
8 m	8.13 (1.22)	8.03 (1.12)	8.33 (1.06)	8.24 (1.21)	7.92 (1.35)
14 m	14.48 (1.27)	14.50 (1.32)	14.56 (1.20)	14.58 (1.29)	14.31 (1.26)
24 m	25.39 (3.06)	24.84 (1.63)	26.40 (4.25)	25.72 (2.31)	24.63 (3.30)
36 m	38.39 (2.32)	38.06 (1.90)	38.19 (2.05)	38.62 (2.29)	38.39 (2.69)
MSEL	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**
GM 8 m	47.24 (10.68)	43.84 (11.37)	45.07 (12.57)	47.31 (10.48)	50.00 (8.68)
GM 14 m	49.57 (14.79)	45.59 (14.46)	47.60 (13.01)	50.87 (14.97)	50.97 (15.54)
GM 24 m	51.54 (11.50)	46.95 (14.02)	49.46 (10.60)	50.91 (11.39)	55.65 (9.72)
FM 8 m	55.10 (12.41)	48.53 (12.91)	52.00 (12.78)	56.77 (12.71)	57.92 (10.17)
FM 14 m	55.69 (9.95)	50.50 (11.63)	52.63 (11.07)	56.50 (8.43)	58.89 (8.83)
FM 24 m	48.77 (10.67)	44.64 (11.55)	45.17 (13.30)	48.49 (8.95)	53.15 (8.83)
FM 36 m	51.40 (16.39)	39.84 (16.09)	43.30 (16.01)	54.34 (14.80)	57.96 (14.02)
VR 8 m	53.81 (11.07)	51.59 (10.46)	50.49 (11.16)	54.35 (11.75)	56.17 (9.95)
VR 14 m	49.99 (9.98)	45.09 (9.31)	48.53 (9.56)	48.95 (9.96)	54.35 (9.05)
VR 24 m	53.62 (12.85)	47.30 (13.08)	46.92 (13.62)	55.18 (10.59)	58.65 (12.14)
VR 36 m	56.83 (13.82)	49.29 (17.54)	49.47 (15.04)	60.51 (10.88)	60.21 (11.29)
RL 8 m	47.56 (10.17)	43.35 (12.50)	46.05 (8.70)	49.90 (10.08)	47.55 (9.33)
RL 14 m	42.64 (11.90)	36.19 (9.05)	40.19 (10.73)	43.68 (12.31)	45.77 (12.01)
RL 24 m	52.17 (12.99)	41.71 (15.47)	46.92 (13.17)	53.26 (10.74)	58.75 (9.71)
RL 36 m	52.53 (12.85)	43.47 (17.71)	43.37 (13.04)	55.80 (8.37)	58.21 (9.15)
EL 8 m	51.05 (10.21)	50.08 (11.89)	51.26 (11.00)	50.52 (10.14)	52.01 (9.07)
EL 14 m	47.77 (10.66)	42.13 (11.44)	44.98 (10.90)	49.64 (9.93)	49.75 (9.93)
EL 24 m	51.60 (12.93)	46.23 (15.30)	47.72 (13.20)	50.70 (11.49)	57.46 (11.18)
EL 36 m	53.95 (12.86)	43.28 (16.14)	45.20 (12.31)	57.76 (9.33)	59.39 (9.38)

TABLE 1 CONTINUED.

	Overall	High-Risk (n = 161)			Low-Risk(n = 71)
		ASD (n = 32)	Atypical (n = 43)	Typical (n = 86)	
VABS	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**
Comm 8 m	95.73 (15.88)	90.19 (15.33)	89.53 (17.12)	96.35 (15.71)	101.23 (13.58)
Comm 14 m	96.59 (13.30)	86.04 (14.28)	93.17 (15.85)	98.35 (11.63)	101.29 (9.68)
Comm 24 m	103.31 (12.73)	94.47 (15.06)	98.41 (10.78)	104.31 (10.66)	109.06 (11.88)
Comm 36 m	101.12 (14.28)	88.96 (18.19)	93.18 (14.03)	103.26 (10.24)	108.82 (10.42)
DL 8 m	99.95 (13.49)	93.56 (15.28)	98.98 (11.75)	101.24 (13.14)	101.87 (13.39)
DL 14 m	95.17 (12.99)	85.63 (13.39)	93.38 (13.38)	96.58 (12.62)	98.86 (10.80)
DL 24 m	105.48 (12.48)	97.88 (13.95)	101.49 (13.50)	107.55 (11.09)	108.83 (10.80)
DL 36 m	103.06 (13.03)	88.36 (18.03)	97.74 (12.64)	106.63 (8.84)	108.60 (7.91)
Mot 8 m	89.65 (16.21)	85.13 (16.71)	80.65 (15.66)	90.73 (15.88)	95.84 (13.80)
Mot 14 m	100.33 (12.84)	98.06 (14.27)	95.73 (15.10)	100.40 (11.59)	104.04 (11.15)
Mot 24 m	99.97 (10.66)	98.19 (12.49)	96.51 (11.83)	99.50 (10.09)	103.44 (8.77)
Mot 36 m	93.66 (12.23)	84.66 (13.25)	86.47 (10.48)	95.79 (10.35)	99.49 (10.54)
Soc 8 m	99.84 (12.72)	96.97 (15.66)	97.93 (11.52)	99.62 (11.85)	102.55 (12.71)
Soc 14 m	97.77 (11.66)	91.40 (11.78)	96.53 (12.66)	98.44 (11.06)	100.58 (10.71)
Soc 24 m	100.72 (11.46)	88.91 (11.68)	97.33 (10.65)	101.99 (8.21)	106.56 (10.74)
Soc 36 m	97.78 (12.89)	79.64 (12.66)	92.43 (11.75)	100.93 (8.68)	105.38 (8.00)
AOSI	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**
8 m	8.34 (4.86)	10.66 (5.71)	9.47 (4.86)	8.51 (4.66)	6.39 (3.96)
14 m	5.10 (4.34)	7.59 (4.42)	7.05 (4.94)	4.48 (3.98)	3.56 (3.41)
ADOS	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**	**mean (SD)**
24 m	2.67 (4.16)	6.59 (7.02)	2.72 (3.19)	2.13 (3.15)	1.52 (2.91)
36 m	6.38 (5.18)	11.19 (6.38)	10.28 (5.92)	3.63 (2.37)	5.17 (3.45)

Note. Detailed statistics on group comparison can be found on *Supplemental Material*. *Abbreviations:* ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; GM= gross motor abilities (MSEL); FM= fine motor abilities (MSEL); VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); VABS = Vineland Adaptive Behavior Scales; Comm = communication skills (VABS); DL = daily living skills (VABS); Soc = social skills (VABS); Mot = motor skills (VABS); AOSI= Autism Observation Scale for Infants; ADOS= Autism Diagnostic Observation Schedule; HR = high-risk siblings; LR = low-risk controls.

LONGITUDINAL CHARACTERIZATION OF DEVELOPMENT

Developmental trajectories were characterised for group contrasts (*LR* vs *HR-Typical* vs *HR-Atypical* vs *HR-ASD*). Figure 1 shows trajectories of MSEL scores, and Figure 2 shows trajectories of VABS scores. Detailed statistics can be found in *Supplemental Material*. Since both MSEL and VABS scores were standard scores normed for age, increasing or decreasing developmental trajectories should be interpreted as individuals developing skills either more rapidly or more slowly than expected based on age-appropriate norms. Moreover, the main effect of age must be examined taking into consideration the interaction with outcome.

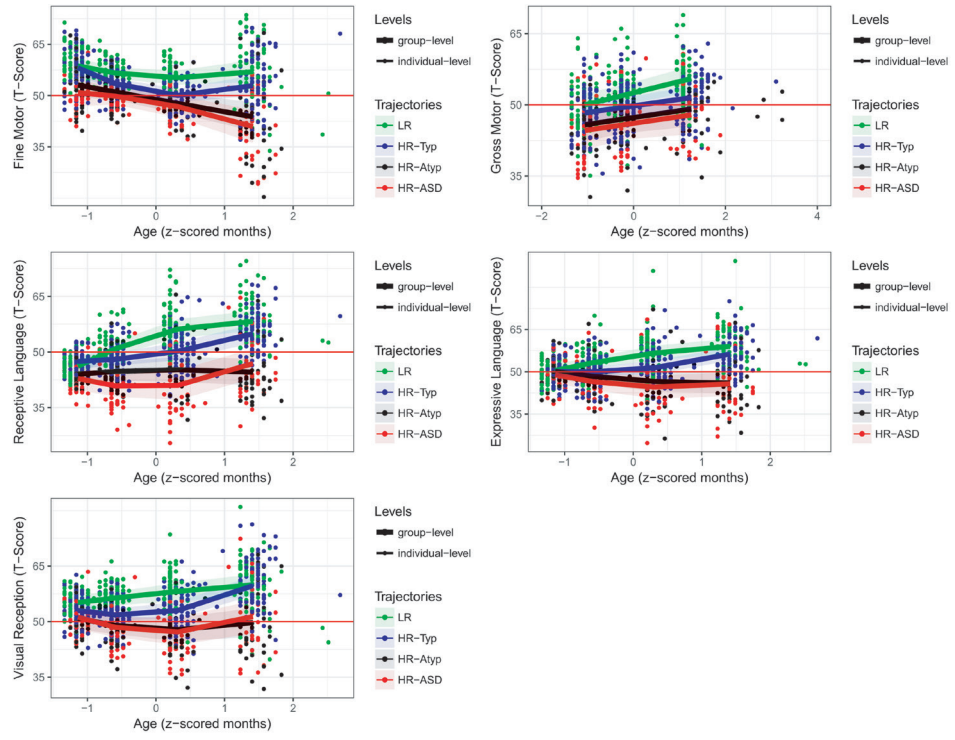


FIGURE 1. Developmental trajectories of estimated means for MSEL measures by clinical outcome groups. This figure shows the longitudinal trajectory of scores per outcome groups (LR, *HR-Typical, HR-Atypical, HR-ASD*) obtained through multilevel mixed modelling for each scale of the MSEL. The developmental trajectories are built on four time-points, one for each visit, which are approximately: 8 months; 14 months; 24 months; 36 months. 95% bootstrap confidence interval on group trajectories is shown as shaded area. Individual scores are also shown (points) with different colours by outcome group. The average normative score is shown by the red line. Abbreviations: MSEL = Mullen Scales of Early Learning; LR = low-risk controls; HR = high-risk siblings.

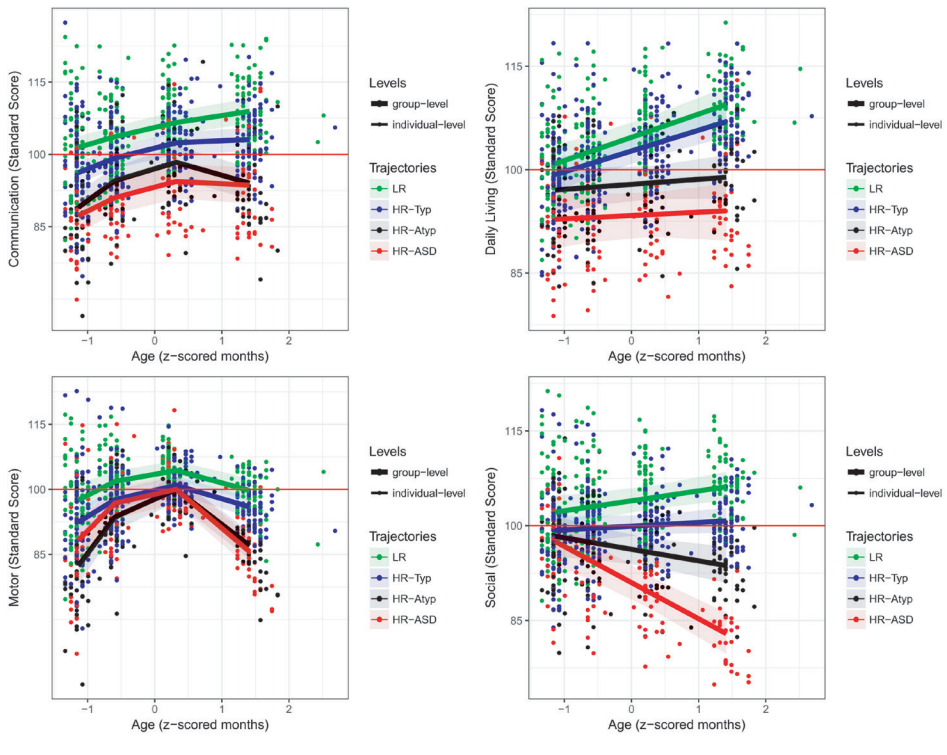


FIGURE 2. Developmental trajectories of estimated means for VABS measures by clinical outcome groups. This figure shows the longitudinal trajectory of scores per outcome groups (LR, *HR-Typical, HR-Atypical, HR-ASD*) obtained through multilevel mixed modelling for each scale of the VABS. The developmental trajectories are built on four time-points, one for each visit, which are approximately: 8 months; 14 months; 24 months; 36 months. 95% bootstrap confidence interval on group trajectories is shown as shaded area. Individual scores are also shown (points) with different colours by outcome group. The average normative score is shown by the red line. Abbreviations: VABS = Vineland Adaptive Behavior Scales; LR = low-risk controls; HR = high-risk siblings.

In terms of trajectories of MSEL scores, we found significant main effect of outcome across all scores (*p*s<0.001, except for GM: p<0.005), and significant main effect of age across all scores (*p*s<0.001) except for EL scores. Specifically, all measures showed quadratic growth (Chi-squared *p*<0.001) except for gross motor scores, which had linear growth over time. Furthermore, the interaction effect between outcome and age was statistically significant for fine motor scores (*p*<0.005), receptive language (*p*<0.005), expressive language (*p*<0.001), and visual reception scores (*p*<0.05). Tukey's post-hoc tests showed that the main effect

of outcome in gross motor scores was mainly driven by a differentiation between *LR* and *HR-Atypical* and *HR-ASD* (respectively *p*=0.018 and *p*=0.007). Similarly, simple main effect analysis showed that the interaction of outcome and age in the other measures was mainly driven by slower increases or decreases in developmental trajectories of *HR-Atypical* and *HR-ASD* compared to LR and *HR-Typical*, leading to an increased differentiation over time which became significant around 14 months. Fine motor scores were the only exceptions, showing significant differentiation between *LR* and *HR-ASD* already by 8 months (*p*<0.05). Across VABS scores, linear growth models were the best fit for daily living and social scores, while communication and motor scores showed quadratic growth. We observed significant main effect of age across all scores (*p*<0.001), and significant main effect of outcome in communication, daily living, and social scores (*p*<0.001). Overall, results showed an increasing gradient of scores from *HR-ASD* to *LR*, with significant differences between *LR* and *HR-ASD* across all scores except for motor scores (Comm and Soc: *p*<0.001; DL: *p*<0.005); between *LR* and *HR-Atypical* across all scores (Comm: *p*<0.005; DL: *p*<0.05; Soc: *p*<0.001; Mot: marginal *p*=0.08); between *HR-Typical* and *HR-Atypical* in daily living (*p*<0.05) and social scores (marginal significance: *p*=0.052); and between *HR-Typical* and *HR-ASD* across all measures except for motor scores (Soc and DL: *p*<0.001; Comm: *p*<0.005). Differences between LR and *HR-Typical* were significant in social scores (*p*<0.05) and marginal in communication scores (*p*=0.07), while *HR-Atypical* only had significantly higher social scores than *HR-ASD* (*p*=0.018). Yet, across all scores the interaction effect between outcome and age was statistically significant (Comm: *outcome x age*^{2}, *p*=0.017; DL: *outcome x age*, *p*=0.024; Soc: *outcome x age*, *p*<0.001; Mot: *outcome x age*^{2}, *p*<0.001). Thus, we performed an analysis of simple main effects. The difference between *LR* and *HR-Typical* and *HR-Atypical* and *HR-ASD* was clear from 8 months in communication, daily living and motor scores. The interaction effect in motor scores was mainly driven by a rapid decrease of scores for *HR-Atypical* and *HR-ASD* between 24 and 36 months, while the initial delay of *HR-Atypical* with respect to *LR* and *HR-Typical* was recovered by 24 months. Similarly, interaction in communication was mainly due to an increase in scores for *LR* and *HR-Typical*, and a decrease for *HR-Atypical* and *HR-ASD* between 24 and 36 months. The interaction in daily living skills was due to an increase of scores over time for *LR* and *HR-Typical*, while *HR-Atypical* and *HR-ASD* were stable below average. Finally, divergent developmental trajectories were clearly visible in social scores: from a steady decrease over time for *HR-ASD* to a slight increase for *LR*, reaching a complete group differentiation at 36 months.

PREDICTING ASD FROM DIFFERENT INSTRUMENTS AND FUNCTIONAL DOMAINS

Next, we classified *HR-ASD* outcome as different from *HR-Atypical* and *HR-Typical* outcome by the integration of data from different instruments and different functional domains at pre-diagnostic ages (i.e. 8 and 14 months), and assessed the added value of data integration when compared to prediction using data from single functional domains or a single instrument. Figure 3 shows the AUC for the different classifiers.

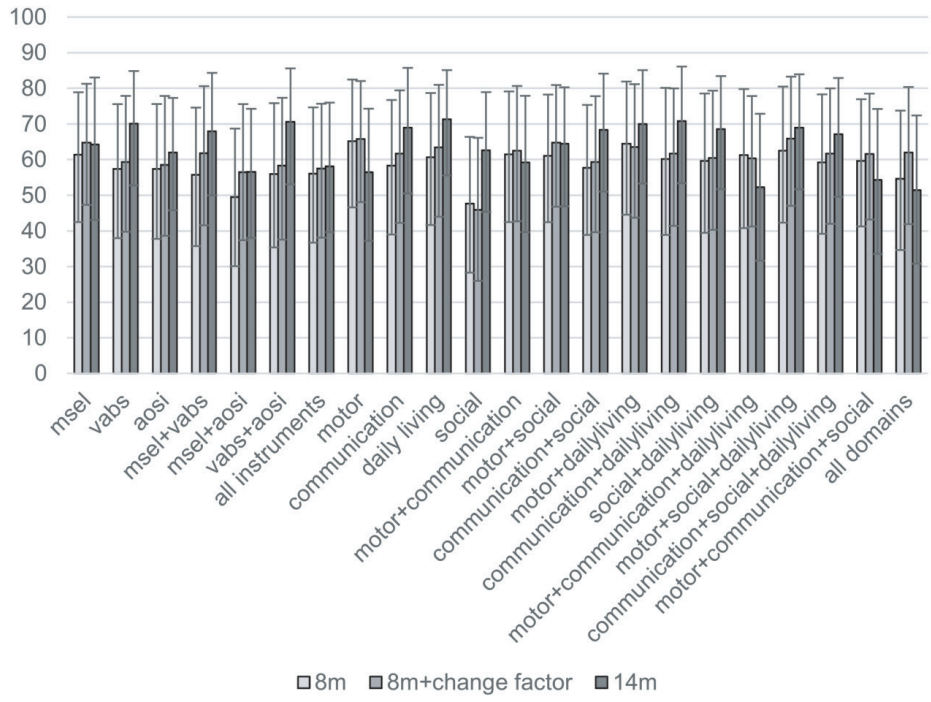


FIGURE 3. Prediction of ASD clinical outcome at 36m: AUC. In this figure the Area Under the Curve (AUC) is reported for different classifiers based on behavioural measures (MSEL, VABS and AOSI) and their combination at different time-points (8 months, 8 months + change factor, 14 months). The classification made is between high-risk infants who are going to develop ASD at 36m, and high-risk infants with typical and atypical (but not ASD) outcome at 36m. The change factor is computed as the difference between measures at 14 and 8 months over the age difference between the 2 visits. The 95% confidence interval is also reported for each classifier. Abbreviations: AUC = area under the curve; MSEL = Mullen Scales of Early Learning (5 scores); VABS = Vineland Adaptive Behavior Scales (4 scores); AOSI = Autism Observation Scale for Infants, in this study we considered the total score.

Considering all classifiers, AUC ranged between 48% and 65% for prediction at 8 months, when all classifiers had predictive performance at chance level. At 14 months, VABS daily living scores showed the best predictive performance (AUC=71.3% [CI, 55.6 to 85.1]; sensitivity=79.6% [CI, 55.2 to 96.6], specificity=52.2% [CI, 38.7 to 65.7], accuracy=57.5% [CI, 45.3 to 69.2], PPV=28.5% [CI, 14.1 to 43.8], NPV=91.5% [CI, 80.3 to 98.7]). This was significantly different from chance level, but not from the classifiers with highest AUC at earlier time points, while differences in performance from other classifiers at 14 months missed significance after Bonferroni correction for multiple comparisons.

To assess the added value of data from multiple time points, we tested predictive performance of classifiers built from cross-sectional scores plus the change factor between time points. As with prediction at 8 months, integrated measures from 8 and 14 months predicted ASD at chance level. However, predictive accuracy improved in comparison to measures at 8 months for the following classifiers: the integration of Mullen and Vineland scores (z=-2.80, p=0.005); communication scores (z=-2.70, p=0.007); the integration of motor, social and daily living scores (z=-2.81, p=0.005); daily living scores (marginally; z=-2.40, p=0.017); and the integration of Vineland scores and the AOSI total score (trend level; z=-2.10, p=0.036). Thus, the rate of developmental change improved predictive accuracy for ASD but prediction was still at chance level. Detailed statistics can be found in *Supplemental Material*.

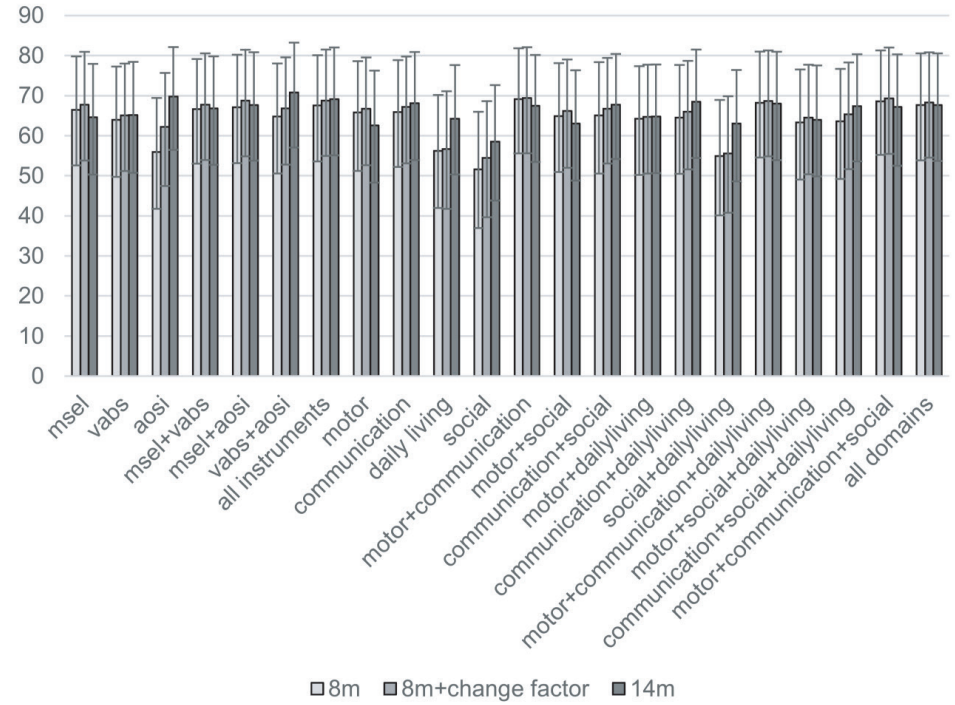


FIGURE 4. Prediction of atypical clinical outcome (including ASD) at 36m: AUC. In this figure the Area Under the Curve (AUC) is reported for different classifiers based on behavioural measures (MSEL, VABS and AOSI) and their combination at different time-points (8 months, 8 months + change factor, 14 months). The classification made is between high-risk infants with atypical development (including an ASD diagnosis at 36m), and high-risk infants with typical outcome at 36m. The change factor is computed as the difference between measures at 14 and 8 months over the age difference between the 2 visits. The 95% confidence interval is also reported for each classifier. Abbreviations: AUC = area under the curve; MSEL = Mullen Scales of Early Learning (5 scores); VABS = Vineland Adaptive Behavior Scales (4 scores); AOSI = Autism Observation Scale for Infants, in this study we considered the total score.

TABLE 2. Best classifiers at each time-point for the two different classifications. Performance metrics for the classifiers chosen as best (based on having the highest AUC) at different age for classifying HR atypically developing siblings (including those who later develop ASD) from their typically developing peers (*HR-ASD + HR-Atypical vs HR-Typical*), and HR sibling who later develop ASD from those who do not (*HR-ASD vs HR-Atypical + HR-Typical*).

Classification	Classifier	p	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	8 months							
HR-ASD *vs. (HR-Atypical +HR-Typical)*	Motor scores	0.11	65.1 (46.6, 82.5)	64.2 (38.1, 89.7)	65.6 (52.7, 78.0)	65.3 (53.6, 76.6)	30.8 (13.6, 49.6)	88.7 (77.9, 97.8)
(HR-ASD + HR-Atypical) *vs. HR-Typical*	Motor + communication scores	0.01*	69.2 (55.6, 81.4)	68.8 (51.6, 85.0)	64.4 (47.8, 79.7)	66.4 (55.2, 77.7)	62.8 (46.0, 79.2)	70.3 (53.5, 85.8)
	8 months + change factor							
HR-ASD *vs. (HR-Atypical +HR-Typical)*	Motor + social + daily living scores	0.10	65.9 (47.0, 83.3)	61.9 (33.4, 88.1)	65.0 (51.5, 77.2)	64.4 (52.5, 75.6)	30.0 (12.7, 47.9)	87.6 (76.1, 96.7)
(HR-ASD + HR-Atypical) *vs. HR-Typical*	Motor + communication scores	0.01*	69.4 (55.6, 82.1)	70.5 (52.9, 86.2)	67.8 (51.7, 82.8)	69.1 (57.5, 80.2)	65.9 (48.7, 81.9)	72.5 (56.4, 87.2)
	14 months							
HR-ASD *vs. (HR-Atypical +HR-Typical)*	Daily living score	0.03*	71.3 (55.6, 85.1)	79.6 (55.2, 96.6)	52.2 (38.7, 65.7)	57.5 (45.3, 69.2)	28.5 (14.1, 43.8)	91.5 (80.3, 98.7)
(HR-ASD + HR-Atypical) *vs. HR-Typical*	VABS scores + AOSI total score	0.01*	70.8 (57.1, 83.2)	60.7 (43.2, 78.3)	67.5 (51.0, 82.7)	64.4 (52.0, 75.9)	62.1 (43.5, 79.5)	66.4 (50.2, 81.6)

Note. The significance of classification AUC was determined by permutation test, the resulting *p-values* are reported. Prediction was considered significant if $p<0.05$ (marked as *). 95% confidence interval is reported in parentheses. All metrics are reported as *mean (lower level CI, upper level CI)*. Abbreviations: AUC = area under the curve; PPV = positive predictive power; NPV = negative predictive power; MSEL = Mullen Scales of Early Learning (5 scores); VABS = Vineland Adaptive Behavior Scales (4 scores); AOSI = Autism Observation Scale for Infants, in this study we considered the total score.

Next, we predicted *HR-ASD* and *HR-Atypical* outcome together as different from *HR-Typical*. The AUC for the different classifiers is shown in Figure 4, while Table 2 shows metrics of the best performing classifiers at each time point for each classification problem. Details on classifier performance at different time points can be found in the *Supplemental Material*.

Considering all classifiers, AUC ranged between 52% and 69% using measures at 8 months, and similarly using measures at 14 months or from integrated time points. By integrating measures from 8 and 14 months, predictive accuracy increased with respect to 8 months

for: AOSI total score (z=-2.70, p=0.007 with α=0.017); the integration of Vineland scores and AOSI total score (z=-2.60, p=0.009); and motor scores (trend level; z=-2.10, p=0.037) (see *Supplemental Material* for details). The best classifier for prediction of *HR-ASD* plus *HR-Atypical* integrated VABS scores and AOSI total scores at 14 months (AUC=71.8% [CI, 58.3 to 83.7]; sensitivity=62.2% [CI, 44.0 to 79.0], specificity=69.5% [CI, 53.1 to 84.6], accuracy=66.1% [CI, 54.4 to 77.0], PPV=63.9% [CI, 45.7 to 81.3], NPV=68.2% [CI, 51.7 to 82.8]). As with classification of ASD, performance of the best classifier at 14 months did not differ from the classifiers with highest AUC at earlier time points, or the other classifiers at 14 months after Bonferroni correction.

Finally, we tested differences in accuracy of classifiers for the two classification problems to examine whether predictive power was different for ASD and broader atypical development and found no statistically significant difference (see *Supplemental Material*).

DISCUSSION

This is the first study in which integrated standardized behavioural measures in infancy were used to characterise developmental trajectories at group-level and to individually classify HR siblings who later develop ASD from their typically and atypically developing peers. Our hypotheses were that integration of information from multiple functional domains and multiple time points would improve early prediction of ASD compared to prediction from a single domain or a single time point. Our main findings were: (1) clear but small size group effects for Mullen and Vineland scores between LR, *HR-ASD*, *HR-Atypical* and *HR-Typical* outcome groups at 8 and 14 months, and larger group effects at 24 and 36 months; (2) individual prediction of ASD from non-ASD outcome at chance level at 8 months, but at moderate and above chance level (AUC=71.3%) at 14 months; (3) individual prediction of broader atypical development from typical outcome with moderate AUC at 8 and 14 months (approximately 70%); (4) added value of combined measures for prediction of broader atypical from typical outcome, but not for prediction of ASD from non-ASD outcome; and (5) added value of combined time points to prediction for some, but not all measures.

DEVELOPMENTAL TRAJECTORIES AT GROUP LEVEL

Differences in development between LR and *HR-Typical* versus *HR-Atypical* and *HR-ASD* drove the differentiation of outcome groups over time. Specifically, developmental trajectories of *LR* and *HR-Typical* infants were either stable or increasing across all scores, showing normative or above average development in respect to age-appropriate norms. In contrast, developmental trajectories of *HR-Atypical* and *HR-ASD* siblings were stable or decreasing across all scores, indicating that those infants tend to fall behind age-appropriate norms during development. This was particularly true for VABS social scores. Furthermore, we observed a gradient of scores across groups in MSEL motor scores and VABS communication, daily living and social scores between 8 and 36 months. Specifically, LR scores were higher than *HR-Typical* scores, which were higher than *HR-Atypical* scores, which were higher than *HR-ASD*. In contrast, VABS motor and MSEL visual reception and language scores showed overlapping or crossing trajectories for *HR-Atypical* and *HR-ASD*. We observed differences between groups from 8 months, supporting and extending results from previous studies which showed differences between *HR-ASD* and *LR*, or *HR-ASD* and *HR-Typical* on several measures at 8 months [46, 47]. In particular, delays in high-risk infants who later developed ASD tended to start in the motor domain at 8 months and extend to the social domain by 14 months. This confirms and extends previous findings to high-risk siblings who received a clinical outcome evaluation at 36 months [14, 35]. However, we found differences between *HR-ASD* and *HR-Typical* on Mullen receptive but not expressive

language scores, and between LR and *HR-ASD* or *HR-Atypical* on Mullen language scores and Vineland communication scores already at 8 months. These differences in communication skills were detected earlier than previously reported [56, 142]. Further work is needed to understand whether the lack of group differentiation on social scores at 8 months is due to the inability of current tools to capture ASD-related manifestations on social skills at early age, or whether it reflects the developmental pathway of ASD. In particular, more granular assessments are needed to characterise developmental trajectories with greater precision and to better capture the dynamics of development.

INDIVIDUAL CLASSIFICATION OF HR-ASD FROM HR NON-ASD.

Next, we moved from group comparisons to individual prediction: our aim was to test whether it was possible to reduce the age at which individual prediction of ASD is possible, and to improve predictive power for ASD using standardized measures. Prior to this study, predictive power of Mullen, Vineland and AOSI scores had not been tested with respect to individual ASD outcome, although these instruments are largely used for clinical evaluation. On the other hand, previous studies have used ADOS scores to classify ASD [113, 114, 143].

To improve predictive power and reduce the age of individual prediction for ASD from behavioural measures, we specifically focused on data from 8 and 14 months, as previous studies classified ASD from ADOS at 18 months [114]. Furthermore, we used different combinations of standardized behavioural measures as predictors, since neuroimaging studies had already shown higher predictive power for ASD from integrated data than data from single modalities [102, 103]. However, our results showed that the integration of different measures did not improve prediction of ASD at 8 months, which remained at chance level. This might be explained by the heterogeneity of the behavioural phenotype linked to later development of ASD in the first year of life, when behavioural atypicalities are subtle and possibly transient. ASD might not be defined as a single category in behaviour before the second year of life, when the defining behaviours generally unfold, explaining poor predictive power of our data.

We also attempted to exploit information from early developmental trajectories as additional information for classification by adding the change factor between 8 and 14 months to cross-sectional scores at 8 months. The rate of development added significant value for prediction to classifiers focused on communication skills; VABS plus MSEL scores; and the integration of motor, social and daily living scores. These results highlight the dynamical changes in development between 8 and 14 months which are relevant to ASD. However, prediction was still not different from chance level, suggesting that prediction of ASD probably depends more on the level of development and functioning rather than

the rate of change. One route to improving classifiers would be to have greater density of data collection between 8 and 14 months to capture developmental dynamics with greater precision. Predictive value was not improved by adding trajectory information to classifiers using information about social skills; the integration of motor and daily living scores; and the integration of motor, communication and daily living scores. For these cases, it is possible that the change factor made the binary separation between classes more difficult by adding heterogeneity due to intra-individual heterogeneity in developmental trajectory. In fact, we found higher intra-individual heterogeneity on fine motor, communication, social, and AOSI scores in *HR-ASD* than other siblings (see *Supplemental Material*).

The Vineland daily living scores at 14 months provided the highest predictive power for ASD (71.3% AUC). Impairments in daily living skills, such as being careful around hot objects or following household rules, are common in children with ASD [144, 145]. However, their investigation is usually underestimated in very young infants because they are difficult to assess at an early age, when parents usually perform tasks for their children. Thus, it is novel to find daily living scores as best predictors for ASD at 14 months. Yet, previous studies have shown that symptom severity in young children with ASD can predict daily living skills [144], and our results show that daily living skills can be affected as soon as (or even before) clinical symptoms begin to emerge. In fact, impairments in daily living skills might reflect the accumulation of more subtle impairments in other domains, like social-communication and motor domains [146, 147], since they require the ability to understand requests and tasks, and to perform the task itself. As a result, more complex actions measured by the daily living skills scale might enlarge the differentiation between infants who later develop ASD and those who do not. However, caution is required because differences in predictive performance from integrated measures (such as communication and daily living scores, motor and daily living scores, or Vineland and AOSI scores) failed to reach statistical significance, and differences from other classifiers were only marginally significant after Bonferroni correction. Nevertheless, it is better in practice to have a simpler predictor and our results suggest that it might be sufficient to assess daily living skills at 14 months.

In summary, despite clear group differences at various levels, individual prediction of ASD using different measures at different time points was still far from optimal. In fact, although the AUC was moderate, our most successful classifier had a much lower PPV than NPV (respectively 28.5% and 91.5%), which means that it was more accurate at predicting infants who will not develop ASD. This might be explained by the prevalence of positives (20% of *HR-ASD*), since low incidence generally reduces PPV, and the measures included in this study, which are tuned to pick up the abilities that define 'typicality'. However, prediction of infants who are going to develop typically in all likelihood is still very useful to allay

any concern. Further work is needed to allow a more accurate prediction of the minority class, for instance including data more specific to ASD. In fact, although AOSI focuses on the assessment of ASD symptoms, behaviours like shyness might be confounding. Thus, moderate classifier performance might link to critical missing variables for prediction of ASD, such as measures of home environment, social attention, or changes in the brain. Furthermore, moderate predictive accuracy might be explained by the high inter-individual variability in clinical symptoms and developmental problems in ASD. Converging evidence suggests the presence of different subgroups within infants who later develop ASD, and the heterogeneity of early developmental pathways to ASD [54, 134]. Measures included in this study might not be able to separate all individuals from different subgroups developing ASD from siblings who do not; other methodological approaches might be used to identify patterns of behaviours predicting ASD specific to the different subgroups. Previous studies already attempted to address heterogeneity by finding separate predictive patterns of symptoms and predicted ASD outcome at 24 months with higher accuracy than the present study [113, 114]. However, we focused on younger age points to predict outcome at 36 months.

INDIVIDUAL CLASSIFICATION OF HR-ASD PLUS HR-ATYPICAL FROM HR-TYPICAL

We used the same classification approach to predict broader atypical development in high-risk siblings. The integration of motor and communication scores from MSEL and VABS allowed classification with moderate accuracy at 8 months (AUC=69.2%), and at combined 8 and 14 months (AUC=69.4%). Differences in AUC between different classifiers (e.g. VABS + AOSI versus VABS alone) suggested that data integration improved predictive performance, though this was only marginally significant after correcting for multiple comparisons. While delays in motor skills have been previously documented in the first year of life [52, 53], the improved predictive accuracy of integrated communication and motor scores is in contrast with previous findings supporting the emergence of ASD in the sensorimotor domain before 12 months, and moving only later to the social-communication domain [14, 35]. However, this might be explained by the inclusion of siblings with early emerging language delays, but not ASD, when classification of ASD is extended to classification of broader atypicality. Nevertheless, it is in line with our results at group level, which showed differentiation between *HR-ASD* and *HR-Typical* on receptive language already at 8 months.

At 14 months, the integration of VABS and AOSI scores showed the highest predictive accuracy (AUC=70.8%), although not statistically significantly different from AOSI alone (AUC=69.8%). Thus, although measures of ASD symptoms seem to retain most of predictive power when taken alone, classifying correctly 70 out of 100 infants, the interplay between symptoms and adaptive functioning improved prediction of broader atypical outcome by

classifying correctly one more infant. The integration of measures from 8 and 14 months also improved predictive accuracy over 8-month data alone for classifiers using AOSI and Vineland plus AOSI scores, in line with previous studies showing an increase in predictive power of AOSI scores in the second year of life [47]. Thus, the rate of emergence of symptoms and the interplay with everyday functioning at the end of the first year of life might be relevant to the development of atypical versus typical outcome. However, it did not provide additional predictive power to measures at 14 months alone.

STRENGTHS AND LIMITATIONS

This study extends previous high-risk studies on early markers for ASD by (1) integrating information from multiple measures and multiple time-points; (2) testing models for individual classification, which is a fundamental issue for clinical practice; (3) focusing on prediction at younger age points. We used a mixed-gender sample for classification and observed a significant difference on gender per clinical outcome, with more males receiving an ASD diagnosis at 36 months than females. Yet, the addition of gender as a feature for classifiers did not significantly improve AUC, except for prediction of ASD from Vineland social scores at 8 months (Bonferroni corrected p=0.02, z=-3.3), and AOSI total scores at 14 months (Bonferroni corrected p=0.01, z=-3.4). Further work should investigate specific differences linked to gender in predictive power for ASD of behavioural and developmental measures. Although we had access to a reasonably large sample, statistical power still remains a limitation to this study. Statistical significance of differences in performance between classifiers with highest AUC and other classifiers at the same time point did not survive Bonferroni correction. Therefore, we need to increase statistical power and results need replication in larger samples. Furthermore, the *HR-Atypical* group needs careful interpretation due to the high variability of individuals included in this group, which was more instrument-defined than clinically based.

FUTURE DIRECTIONS

Despite clear differences at group level and moderately high predictive accuracy, individual prediction still needs to be improved. Our results provide further evidence to the high inter-individual and intra-individual heterogeneity of ASD, which makes difficult to predict the later development of the disorder at an early age. Further investigation is needed to understand the interplay of different domains in the first years of life leading to an ASD outcome. It is also possible that the ASD behavioural phenotype simply does not exist as a definable category before two years of age. The exploratory investigation of heterogeneity of ASD development in a bottom-up approach through latent class analysis can provide better insight into the different developmental trajectories leading to ASD, and a data driven approach might be used to discover new categories than those currently used for

classification. Furthermore, the combination of measures from different domains can be extended to include more biological data (e.g. genetics and epigenetics, EEG and ERP, or functional MRI data) and measures of environmental experience (e.g. parent behaviour or socioeconomic status) to provide a more complete picture of the developmental status of the infant. Future work must also investigate generalizability of predictive classifiers to provide a useful tool for clinical practice.

SUPPLEMENTARY MATERIAL

MISSING DATA

TABLE S1. Missing data. This table shows in the second column the number of infants attending each visit (*n/n_total*), where *n_total*=237 is the total number of infants after excluding infants who did not receive a clinical evaluation and/or an ADOS classification at 36 months. In the third column, the number of infants with all complete scores is shown over the number of participants for each visit (*n_attendance*).

Visit	Attendance (n/n_total)	Complete subjects (n/n_attendance)
8 months	237/237	231/237
14 months	234/237	223/234
24 months	235/237	158/235
36 months	237/237	230/237

Data presented in the current paper were collected as part of a large longitudinal study, to which 247 infants participated in one of two phases of longitudinal assessments (104 in Phase 1 and 143 in Phase 2). Missing data was mainly due to non-attendance to visits. N=10 infants were excluded from this study because they did not receive an ADOS (Autism Diagnostic Observation Schedule) evaluation and/or a clinical outcome evaluation at 36 months. We investigated the pattern of missing data in the selected sample (N=237) at different time-points, testing differences on risk-group (high-risk siblings, HR; low-risk controls, LR), gender, age and clinical outcome at 36 months (LR; *HR-Typical*; *HR-Atypical*; *HR-ASD*) between infants with complete and missing data. Table S1 shows the number of infants with complete data among those who attended the visit. We found a total of 3.5% of data missing, and differences on risk-group ($p<0.001$, $t=-8.35$, $df=233$), gender ($p=0.04$, $t=2.05$, $df=233$), and clinical outcome ($p<0.001$, $t=-6.62$, $df=233$) were significant at 24 months, showing a pattern of data missing at random. We performed imputation through expectation maximization to handle missing data at a specific time-point. Analyses were performed on SPSS (http://www.ibm.com/analytics/us/en/technology/spss). Our aim was to obtain a longitudinally complete dataset for each infant between 8 and 36 months, thus infants who did not attend at least one of the visits were excluded from the study (N=5). Our final sample included 232 infants (161 [69.4%] HR and 71 [30.6%] LR). Infants excluded from the study (N=15) did not differ from the selected sample on risk group ($p=0.45$, $t=0.76$, $df=245$), gender ($p=0.21$, $t=-1.25$, $df=245$), and age at intake ($p=0.53$, $t=-0.63$, $df=245$), showing a pattern of data missing completely at random.

ATYPICAL CLINICAL OUTCOME

Criteria for the atypical outcome were: ADOS above ASD (Autism Spectrum Disorder) threshold (N=31) or ADI-R (Autism Diagnostic Interview – Revised) above ASD threshold (N=6) or MSEL (Mullen Scales of Early Learning) more than 1.5 standard deviations below the average on visual reception (N=7), receptive language (N=13), expressive language (N=9), or early learning composite score (N=15). Among these, N=20 infants met only the ADOS criterion; N=11 met only the MSEL criterion; N=1 met only the ADI-R criterion; N=6 met ADOS and MSEL criteria; N=4 met ADOS and ADI-R criteria; and N=1 met all criteria.

INTRA-INDIVIDUAL VARIABILITY

Pearson correlation between measures at 8 and 14 months by clinical outcome groups were computed as a measure of intra-individual variability in developmental trajectories between the time-point used for classification. Analyses were performed on Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016). Results are shown in Table S2.

Table S2. Intra-individual variability between 8 and 14 months. This table reports the Pearson correlation coefficient between single measures at 8 and 14 months by clinical outcome groups (LR, HR-Typical, HR-Atypical, HR-ASD). The number of infants in each group is also reported (*n*).

Measure	HR-ASD (n=32)	HR-Atypical (n=43)	HR-Typical (n=86)
GM	0.67	0.23	0.52
VR	0.40	0.44	0.26
FM	0.11	0.50	0.34
RL	0.40	-0.02	0.11
EL	0.43	0.34	0.17
Comm	0.18	0.37	0.40
DL	0.40	0.38	0.24
Soc	0.15	0.59	0.43
Mot	0.74	0.61	0.52
AOSI total score	0.15	0.35	0.29

Abbreviations: HR= high-risk siblings; LR= low-risk controls; ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; GM= gross motor abilities (MSEL); FM= fine motor abilities (MSEL); VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); VABS = Vineland Adaptive Behavior Scales; Comm = communication skills (VABS); DL = daily living skills (VABS); Soc = social skills (VABS); Mot = motor skills (VABS); AOSI= Autism Observation Scale for Infants.

DEVELOPMENTAL TRAJECTORIES BY RISK GROUPS

Figures S1 and S2 show developmental trajectories by risk status (HR vs. LR) respectively in the Mullen and Vineland scales between 8 and 36 months. LR controls showed higher scores than HR siblings in all measures of developmental level and adaptive functioning throughout development. Trajectories were obtained by multilevel mixed modelling and detailed statistics for each measure are reported below.

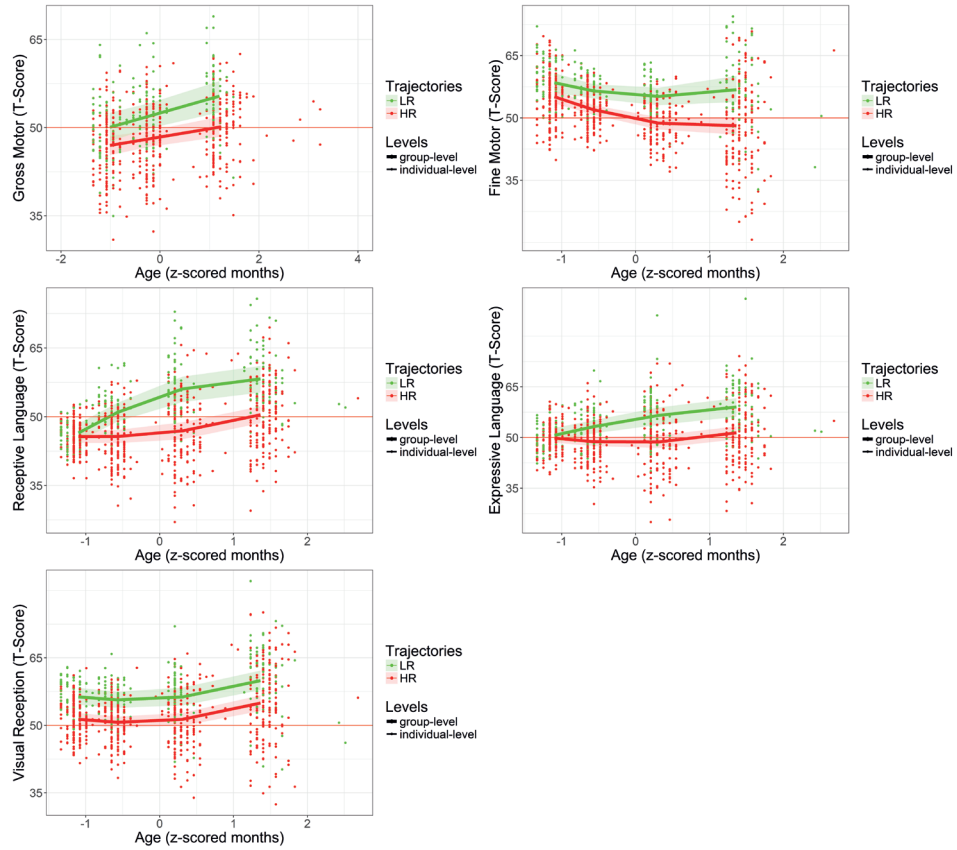


FIGURE S1. Developmental trajectories of estimated means for MSEL measures by risk groups. This figure shows the longitudinal trajectory of scores per risk groups (LR, HR) obtained through multilevel mixed modelling for each scale of the MSEL. The developmental trajectories are built on four time-points, one for each visit, which are approximately: 8 months; 14 months; 24 months; 36 months. 95% bootstrap confidence interval on group trajectories is shown as shaded area. Individual scores are also shown (points) with different colours by group. The average normative score is shown by the red line. *Abbreviations:* MSEL = Mullen Scales of Early Learning; LR = low-risk controls; HR = high-risk siblings.

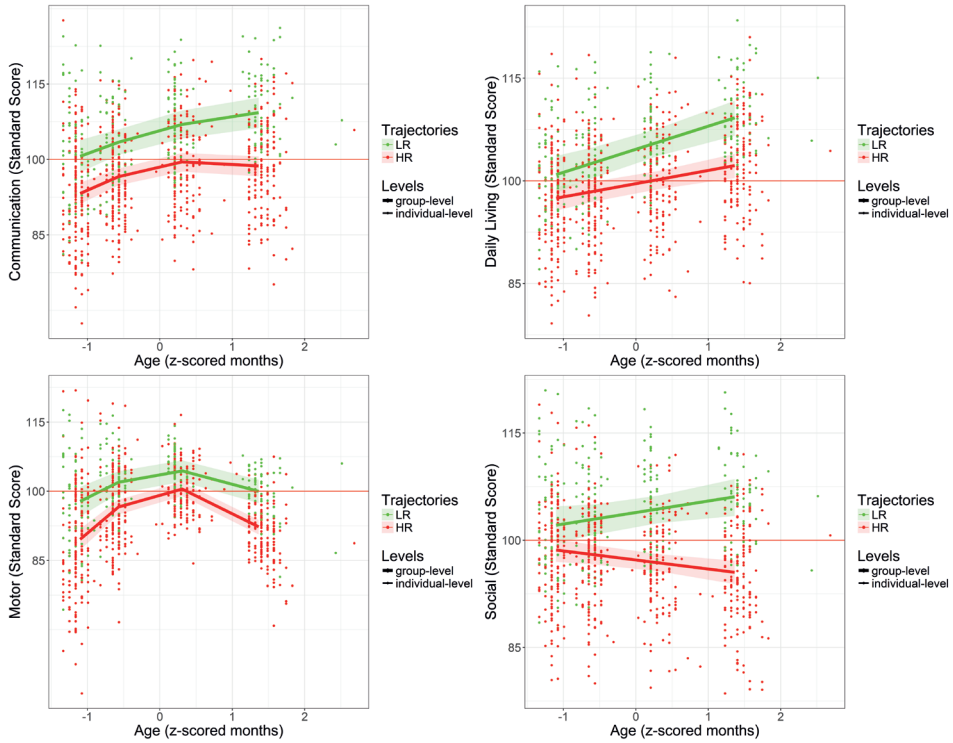


FIGURE S2. Developmental trajectories of estimated means for VABS measures by risk groups. This figure shows the longitudinal trajectory of scores per risk groups (LR, HR) obtained through multilevel mixed modelling for each scale of the VABS. The developmental trajectories are built on four time-points, one for each visit, which are approximately: 8 months; 14 months; 24 months; 36 months. 95% bootstrap confidence interval on group trajectories is shown as shaded area. Individual scores are also shown (points) with different colours by group. The average normative score is shown by the red line. *Abbreviations:* VABS = Vineland Adaptive Behavior Scales; LR = low-risk controls; HR = high-risk siblings.

Gross motor score

The quadratic model was unidentifiable having only 3 time points available for measurements, thus we chose the linear model for data modeling. Furthermore, the model with interaction between outcome and age was not significantly better than the model without interaction effects ($\chi^2(1)=1.3$, p=0.25), showing no interaction effect between age and outcome for gross motor scores. We found a significant main effect of age ($F(1,475.5)=19.2$, $p<0.001$), and risk status ($F(1, 229.3)=9.8$, $p=0.002$) on Mullen gross motor scores. Post-hoc group comparisons showed LR having higher scores than HR siblings (p=0.002).

Fine motor score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=38.3$, p<0.001), and similarly the model with interaction between risk status and age was better than the model without interaction effects ($\chi^2(5)=31.3$, p<0.001). We found a significant main effect of age ($F(1,226.8)=23.6$, $p<0.001$), age^2 ($F(1,291.7)=11.0$, $p=0.001$) and risk status ($F(1, 304.5)=21.9$, $p<0.001$) on Mullen fine motor scores. There was also a significant effect of gender covariate ($F(1,240.8)=12.7$, $p<0.001$) and an interaction effect of age with risk status ($F(1,226.8)=5.8$, $p=0.02$). Post-hoc group comparisons on simple main effects showed LR having higher scores than HR at 14, 24 and 36 months (*p*s<0.001) with the trajectory for HR siblings decreasing significantly between 8 and 14 months, and 14 and 24 months (*p*s<0.001).

Receptive language score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=18.7$, p=0.002), and similarly the model with interaction between risk status and age was better than the model without interaction effects ($\chi^2(5)=30.6$, p<0.001). We found a significant main effect of age ($F(1,226.1)=66.8$, $p<0.001$) and risk status ($F(1, 261.7)=27.9$, $p<0.001$) on Mullen receptive language scores. There was also a significant effect of gender covariate ($F(1,231.9)=14.7$, $p<0.001$) and an interaction effect of age ($F(1,225.9)=18.1$, $p<0.001$) and age^2 with risk status ($F(1,704.6)=11.1$, $p<0.001$). Post-hoc group comparisons on simple main effects showed LR having higher scores than HR at 14, 24 and 36 months (*p*s<0.001). Furthermore, LR controls showed increasing skills between 8 and 14 months (p<0.001) and 14 and 24 months (p<0.001), while HR siblings showed increasing receptive language skills only between 24 and 36 months (p=0.001).

Expressive language score.

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=37.8$, p<0.001), and similarly the model with interaction between risk status and age was better than the model without interaction effects ($\chi^2(5)=43.7$, p<0.001). We found a significant main effect of age ($F(1, 226.2)=16.7$, $p<0.001$), risk status ($F(1, 241.5)=20.5$, $p<0.001$) and gender covariate ($F(1,230.7)=6.6$, $p=0.01$). Furthermore, we found an interaction effect of age ($F(1,226.1)=12.4$, $p<0.001$) and age^2 with risk status ($F(1,557.4)=7.1$, $p=0.008$). Post-hoc group comparisons on simple main effects showed LR having higher scores than HR at 14 (p=0.005), 24 and 36 months (*p*s<0.001), with a significant increase in scores for LR between 14 and 24 months (p<0.001) and for HR between 24 and 36 months (p=0.01).

Visual reception score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=19.9$, p=0.001), but the model with interaction between risk status and age was not significantly better than the model without interaction effects ($\chi^2(5)=8.3$, p=0.14). We found a significant main effect of age ($F(1,287.7)=7.0$, $p=0.009$), age^2 ($F(1,571.3)=12.0$, $p<0.001$) and risk status ($F(1, 226.3)=22.7$, $p<0.001$). There was also a significant effect of gender covariate ($F(1,228.2)=13.6$, $p<0.001$). Post-hoc group comparisons showed LR having higher scores than HR siblings throughout development (p<0.001).

Communication score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=40.0$, p<0.001), but the model with the interaction between risk status and age was not significantly better than the model without interaction effects ($\chi^2(1)=5.2$, p=0.07). We found a significant main effect of age ($F(1,220.9)=36.7$, $p<0.001$), age^2 ($F(1,215.8)=9.6$, $p=0.002$) and risk status ($F(1, 234.4)=18.7$, $p<0.001$) on Vineland communication scores. There was also a significant effect of gender covariate ($F(1,229.3)=20.4$, $p<0.001$). Post-hoc group comparisons showed LR having higher scores than HR throughout development (p<0.001).

Daily living score

Quadratic model fitting was not significantly better than linear fitting ($\chi^2(5)=8.6$, p=0.13), thus the linear model was selected for model fitting. Furthermore, the model with interaction between risk status and age was not significantly better than the model without interaction effects ($\chi^2(1)=2.9$, p=0.09). We found a significant main effect of age ($F(1,218.6)=39.1$, $p<0.001$), and risk status ($F(1, 229.1)=15.5$, $p<0.001$) on Vineland daily living scores. There was also a significant effect of gender covariate ($F(1,229.2)=14.6$, $p<0.001$). Post-hoc group comparisons showed LR having higher scores than HR siblings throughout development (p<0.001).

Social score

Quadratic model fitting was not significantly better than linear fitting ($\chi^2(5)=7.7$, p=0.18), thus the linear model was selected for model fitting. Furthermore, the model with interaction between risk status and age was better than the model without interaction effects ($\chi^2(1)=235.6$, p<0.001). The main effect of age was not significant, but we found a significant main effect of risk status ($F(1,229.0)=33.0$, $p<0.001$) and gender covariate ($F(1,229.0)=16.3$, $p<0.001$), and a significant interaction effect of age with risk status ($F(1,225)=11.9$, $p<0.001$). Post-hoc group comparisons on simple main effects showed LR having higher scores than HR siblings at 14 months (p=0.002), and 24 and 36 months (p<0.001).

Motor score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(5)=163.2$, p<0.001), and similarly the model with interaction between risk status and age was better than the model without interaction effects ($\chi^2(5)=109.1$, p<0.001). We found a significant main effect of age ($F(1,224.1)=21.4$, $p<0.001$), age^2 ($F(1,198.4)=109.1$, $p<0.001$), and risk status ($F(1, 219.6)=7.9$, $p=0.006$) on Vineland motor scores. There was also a significant effect of gender covariate ($F(1,226.3)=7.9$, $p=0.005$) and an interaction effect of age^2 with risk status ($F(1,198.3)=8.1$, $p=0.005$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Atypical* at 8 (p<0.001),14 (p<0.05), and 36 months (p<0.001), and higher scores than HR siblings at 8 months (p=0.002), 14 months (p=0.01) and 36 months (p<0.001). Furthermore, LR controls showed increasing motor scores between 8 and 14 months (p=0.001), while HR siblings showed increasing scores between 8 and 14 months (p<0.001), and 14 and 24 months (p<0.001).

DEVELOPMENTAL TRAJECTORIES BY CLINICAL OUTCOME GROUPS

Developmental trajectories by clinical outcome groups were shown and discussed in the main text. Details on statistics are reported below.

Gross motor score

The quadratic model was unidentifiable having only 3 time points available for measurements, thus we chose the linear model for data modeling. Furthermore, the model with interaction between outcome and age was not significantly better than the model without interaction effects ($\chi^2(3)=1.3$, p=0.72), showing no interaction effect between age and outcome for gross motor scores. We found a significant main effect of age ($F(1,471.2)=15.5$, $p<0.001$), and outcome ($F(3, 227.2)=4.8$, $p<0.005$) on Mullen gross motor scores. Post-hoc group comparisons showed LR having higher scores than *HR-Atypical* (p=0.018) and *HR-ASD* (p=0.007).

Fine motor score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(7)=42.7$, p<0.001), and similarly the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(9)=41.2$, p<0.001). We found a significant main effect of age ($F(1,227.6)=40.0$, $p<0.001$), age^2 ($F(1,330.4)=4.00$, $p<0.05$) and outcome ($F(3, 304.7)=8.2$, $p<0.001$) on Mullen fine motor scores. There was also a significant effect of gender covariate ($F(1,234.7)=9.2$, $p<0.005$) and an interaction effect of age^2 with outcome ($F(3,324.0)=4.0$, $p<0.005$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-ASD* at all time-points (p<0.05 at 8 months, p<0.005 at 14 and 24 months, p<0.001 at 36 months), higher scores than *HR-Typical* at 24 months (p<0.05),

and higher scores than *HR-Atypical* at 14 (p<0.05), 24 (p<0.005) and 36 months (p<0.001). Furthermore, *HR-Typical* had higher scores than *HR-Atypical* (trend level, p=0.07) and *HR-ASD* (p<0.05) at 36 months.

Receptive language score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(7)=23.5$, p<0.005), and similarly the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(9)=44.4$, p<0.001). We found a significant main effect of age ($F(1,229.1)=30.0$, $p<0.001$) and outcome ($F(3, 272)=15.1$, $p<0.001$) on Mullen receptive language scores. There was also a significant effect of gender covariate ($F(1,231.8)=9.4$, $p<0.005$) and an interaction effect of age ($F(3,228.7)=7.8$, $p<0.001$) and age^2 with outcome ($F(3,588.4)=5.0$, $p<0.005$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Atypical* at 14 (p<0.05), 24 and 36 months (p<0.001), and higher scores than *HR-ASD* from 14 months onwards (p<0.001). Furthermore, *HR-Typical* had higher scores than *HR-Atypical* at 36 months (p<0.001), and higher scores than *HR-ASD* from 14 months onwards (p<0.05).

Expressive language score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(7)=40.0$, p<0.001), and similarly the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(9)=66.2$, p<0.001). We did not find a significant main effect of age or age^2, but a significant main effect of outcome ($F(3, 244.8)=9.6$, $p<0.001$) and gender covariate ($F(1,227.4)=4.1$, $p<0.05$). Furthermore, we found an interaction effect of age with outcome ($F(3,227.1)=9.1$, $p<0.001$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Atypical* at 24 (p<0.005) and 36 months (p<0.001), and higher scores than *HR-ASD* at 14 months (p<0.05), 24 and 36 months (p<0.001). Furthermore, *HR-Typical* had higher scores than *HR-Atypical* and *HR-ASD* at 36 months (p<0.005).

Visual reception score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(7)=20.3$, p=0.005), and similarly the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(9)=19.0$, p=0.03). We found a significant main effect of age^2 ($F(1,235.0)=11.9$, $p<0.001$) and outcome ($F(3, 225.0)=12.7$, $p<0.001$). There was also a significant effect of gender covariate ($F(1,226.7)=10.2$, $p<0.005$) and an interaction effect of age ($F(3,220.8)=3.1$, $p=0.03$) and age^2 with outcome ($F(3,218.9)=2.7$, $p=0.05$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Typical* at 14 months (p<0.05), higher scores than *HR-Atypical* from 14 months onwards (p<0.001),

and higher scores than *HR-ASD* at 14 and 24 months (p<0.001) and 36 months (p<0.05). Furthermore, *HR-Typical* at 36 months had higher scores than *HR-Atypical* (p<0.005) and *HR-ASD* (p<0.05).

Communication score

Quadratic model fitting was significantly better than linear fitting ($\chi^2(7)=43.3$, p<0.001), and similarly the model with the interaction between outcome and age^2 was better than the model without interaction effects ($\chi^2(3)=10.2$, p=0.017). We found a significant main effect of age ($F(1,222.7)=38.4$, $p<0.001$), age^2 ($F(1,274.6)=13.8$, $p<0.001$) and outcome ($F(3, 242.4)=22.4$, $p<0.001$) on Vineland communication scores. There was also a significant effect of gender covariate ($F(1,227.2)=13.3$, $p<0.001$) and an interaction effect of age^2 with outcome ($F(3,195.2)=3.5$, $p=0.017$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Atypical* and *HR-ASD* at all time-points (p<0.001, except for *HR-Atypical* at 24 months with p<0.005), and higher scores than *HR-Typical* at 36 months (p<0.05); and *HR-Typical* having higher scores than *HR-Atypical* at 8 (p<0.05) and 36 months (p<0.005), and higher scores than *HR-ASD* at all time-points ($p<0.001$ at 8 months, p<0.05 at 14 and 24 months, and p<0.005 at 36 months).

Daily living score

Quadratic model fitting was not significantly better than linear fitting ($\chi^2(7)=9.7$, p=0.21), thus the linear model was selected for model fitting. Furthermore, the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(3)=9.6$, p=0.02). We found a significant main effect of age ($F(1,227.4)=20.5$, $p<0.001$), and outcome ($F(3, 227)=15.9$, $p<0.001$) on Vineland daily living scores. There was also a significant effect of gender covariate ($F(1,226.9)=9.0$, $p<0.005$) and an interaction effect of age with outcome ($F(3,225.2)=3.2$, $p=0.02$). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-ASD* at all time-points (p<0.001), and higher scores than *HR-Atypical* at 24 and 36 months (p<0.001); and *HR-Typical* having higher scores than *HR-Atypical* at 24 and 36 months (p<0.05), and higher scores than *HR-ASD* at 14 ($p<0.05$), 24 and 36 months (p<0.001).

Social score

Quadratic model fitting was not significantly better than linear fitting ($\chi^2(7)=11.5$, p=0.12), thus the linear model was selected for model fitting. Furthermore, the model with interaction between outcome and age was better than the model without interaction effects ($\chi^2(3)=273.5$, p<0.001). We found a significant main effect of age ($F(1,226.9)=11.9$, $p<0.001$), and outcome ($F(3, 227.2)=23.2$, $p<0.001$) on Vineland social scores. There was also a significant effect of gender covariate ($F(1,226.3)=9.7$, $p<0.005$) and an interaction effect of

age with outcome (F(3,225.5)=14.7, p<0.001). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Typical* at 24 and 36 months (p<0.05), *HR-Atypical* at 14 (p<0.05), 24 and 36 months (p<0.001), and *HR-ASD* from 14 months onwards (p<0.001); *HR-Atypical* having higher scores than *HR-ASD* at 24 (p<0.05) and 36 months (p<0.001); and *HR-Typical* having higher scores than *HR-Atypical* at 36 months (p<0.05), and *HR-ASD* at 24 and 36 months (p<0.001).

Motor score

Quadratic model fitting was significantly better than linear fitting (χ^2(7)=183.4, p<0.001), and similarly the model with interaction between outcome and age was better than the model without interaction effects (χ^2(9)=121.2, p<0.001). We found a significant main effect of age (F(1,223.7)=25.6, p<0.001), age^2 (F(1,242.9)=187.1, p<0.001), but only marginally significant main effect of outcome (F(3, 225.5)=2.6, p=0.053) on Vineland motor scores. There was also a significant effect of gender covariate (F(1,224.4)=6.1, p=0.014) and an interaction effect of age^2 with outcome (F(3,223.2)=10.6, p<0.001). Post-hoc group comparisons on simple main effects showed LR having higher scores than *HR-Atypical* at 8 (p<0.001),14 (p<0.05), and 36 months (p<0.001), and higher scores than *HR-ASD* at 36 months (p<0.001); and *HR-Typical* having higher scores than *HR-Atypical* (p<0.05) and *HR-ASD* at 36 months (p<0.005).

CLASSIFIERS

To predict autism at pre-diagnostic ages, we performed a classifier analysis using scores from MSEL, VABS and AOSI as features. Twenty-two classifiers were tested, and features for each classifier are shown in Table S3.

TABLE S3. Classifiers. This table describes the feature composition of classifiers used to predict ASD and atypical outcome at 36 months.

Classifier	Features
msel	GM + FM + RL + EL + VR
vabs	Comm + DL + Mot + Soc
aosi	AOSI total score
msel + vabs	GM + FM + RL + EL + VR + Comm + DL + Mot + Soc
msel + aosi	GM + FM + RL + EL + VR + AOSI total score
vabs + aosi	Comm + DL + Mot + Soc + AOSI total score
all instruments	GM + FM + RL + EL + VR + Comm + DL + Mot + Soc + AOSI total score
motor	GM + FM + Mot
communication	RL + EL + Comm
daily living	DL
social	Soc
motor + communication	GM + FM + Mot + RL + EL + Comm
motor + social	GM + FM + Mot + Soc
communication + social	RL + EL + Comm + Soc
motor + daily living	GM + FM + Mot + DL
communication + daily living	RL + EL + Comm + DL
social + daily living	Soc + DL
motor + communication + daily living	GM + FM + Mot + RL + EL + Comm + DL
motor + social + daily living	GM + FM + Mot + Soc + DL
communication + social + daily living	RL + EL + Comm + Soc + DL
motor + communication + social	GM + FM + Mot + RL + EL + Comm + Soc
all domains	GM + FM + Mot + RL + EL + Comm + Soc + DL

Abbreviations: ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; GM= gross motor abilities (MSEL); FM= fine motor abilities (MSEL); VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); VABS = Vineland Adaptive Behavior Scales; Comm = communication skills (VABS); DL = daily living skills (VABS); Soc = social skills (VABS); Mot = motor skills (VABS); AOSI= Autism Observation Scale for Infants.

PREDICTION OF HR-ASD VS. HR-ATYPICAL AND HR-TYPICAL

Table S4 shows performance measured by Area Under the Curve (AUC) of different classifiers at predicting *HR-ASD* vs. other high-risk siblings. Classification before 14 months was not significantly different from chance level, thus we did not perform comparative analyses on the performance of those classifiers. Using measures at 14 months, prediction

was significantly different from random, thus we performed a nonparametric Friedman test on classifier performance (AUC) at 14 months. We found a significant difference in classifier performance, $\chi^2(21)=117$, $p<0.001$; yet post-hoc comparisons through paired Wilcoxon test missed significance after Bonferroni correction for multiple comparison ($\alpha_{Bonferroni}=0.0024$; see Table S5). The classifier with highest AUC was built on daily living scores at 14 month, yet paired Wilcoxon tests between this classifier and the ones with highest AUC at 8 months (motor scores) and 8 months plus the change factor between 8 and 14 months (motor + social + daily living scores) missed significance (respectively $z=-1.9$, $p=0.06$; $z=-1.7$, $p=0.09$). We also tested significant changes on performance of the same classifier over time by means of a nonparametric Friedman test, and Bonferroni corrected post-hoc paired Wilcoxon tests (3 pairs, $\alpha_{Bonferroni}=0.017$). Results are shown in Table S6.

TABLE S4. Classifier performances for prediction of ASD outcome at 36 months. Predictive performance as measured by the AUC of the classifiers using as features developmental, behavioural and symptoms measures for classifying HR siblings who later develop ASD from their typically developing and atypical non-ASD peers (*HR-ASD vs HR-Typical + HR-Atypical*).

Classifier	8 months		8 months + change factor		14 months	
	p	AUC (%)	*p*	AUC (%)	*p*	AUC (%)
msel	0.21	61.3 (42.5, 78.9)	0.13	64.8 (47.3, 81.3)	0.10	64.2 (43.1, 83.0)
vabs	0.31	57.4 (38.0, 75.6)	0.23	59.4 (39.8, 77.9)	0.047*	70.1 (52.7, 84.9)
aosi	0.25	57.4 (37.8, 18.2)	0.22	58.5 (38.6, 77.9)	0.13	62.0 (45.8, 77.3)
msel + vabs	0.32	55.7 (35.7, 74.6)	0.17	61.8 (41.6, 80.6)	0.07	67.9 (50.0, 84.3)
msel + aosi	0.54	49.5 (30.1, 68.7)	0.32	56.5 (37.4, 75.6)	0.34	56.6 (38.0, 74.3)
vabs + aosi	0.30	56.0 (35.4, 75.8)	0.24	58.3 (37.5, 77.4)	0.04*	70.6 (53.1, 85.6)
all instruments	0.28	56.0 (36.7, 74.7)	0.25	57.5 (38.1, 75.7)	0.32	58.1 (39.7, 76.0)
motor	0.11	65.1 (46.6, 82.5)	0.12	65.8 (48.1, 82.1)	0.30	56.5 (37.2, 74.3)
communication	0.24	58.3 (39.0, 76.7)	0.17	61.7 (42.2, 79.4)	0.03*	69.0 (50.5, 85.7)
daily living	0.21	60.7 (41.6, 78.7)	0.14	63.4 (44.0, 81.0)	0.03*	71.3 (55.6, 85.1)
social	0.58	47.6 (28.3, 66.4)	0.64	45.9 (26.0, 66.1)	0.14	62.6 (45.4, 78.9)

TABLE S4 CONTINUED.

Classifier	8 months		8 months + change factor		14 months	
	p	AUC (%)	*p*	AUC (%)	*p*	AUC (%)
motor + communication	0.17	61.5 (42.5, 79.1)	0.16	62.5 (42.7, 80.7)	0.27	59.2 (39.7, 77.9)
motor + social	0.19	61.0 (42.4, 78.3)	0.10	64.7 (46.8, 80.9)	0.12	64.4 (46.9, 80.3)
communication + social	0.27	57.7 (38.8, 75.3)	0.22	59.3 (39.6, 77.8)	0.047*	68.3 (51.0, 84.1)
motor + daily living	0.14	64.4 (44.5, 81.9)	0.18	63.6 (43.7, 81.2)	0.04*	70.0 (53.3, 85.1)
communication + daily living	0.21	60.2 (38.9, 80.1)	0.18	61.6 (41.4, 80.0)	0.02*	70.8 (53.4, 86.1)
social + daily living	0.26	59.6 (39.5, 78.6)	0.21	60.5 (40.3, 79.3)	0.07	68.6 (51.7, 83.4)
motor + communication + daily living	0.17	61.2 (40.8, 79.8)	0.22	60.4 (41.2, 77.8)	0.41	52.3 (31.7, 72.9)
motor + social + daily living	0.17	62.5 (42.3, 80.5)	0.10	65.9 (47.0, 83.3)	0.06	68.9 (51.7, 83.9)
communication + social + daily living	0.25	59.2 (39.2, 78.3)	0.20	61.7 (42.0, 80.0)	0.09	67.1 (49.5, 82.9)
motor + communication + social	0.24	59.6 (41.2, 76.9)	0.16	61.6 (43.2, 78.5)	0.37	54.3 (33.5, 74.2)
all domains	0.38	54.6 (34.6, 73.8)	0.15	62.0 (41.9, 80.4)	0.46	51.5 (30.8, 72.4)

Notes. The significance of classification AUC was determined by permutation test, the resulting p-values are reported. Prediction was considered different from chance level if $p<0.05$ (marked as *). 95% confidence interval is reported in parentheses. Measures are reported as *mean (lower level CI, upper level CI). Abbreviations:* AUC = area under the curve; MSEL = Mullen Scales of Early Learning (5 scores); VABS = Vineland Adaptive Behavior Scales (4 scores); AOSI = Autism Observation Scale for Infants, in this study we considered the total score.

TABLE S5. Paired Wilcoxon tests on classifier performance for ASD prediction at 14 months. This table shows results from paired Wilcoxon tests (*z-score* and two-tailed *p-value*) on predictive performance, measured by Area Under the Curve (AUC), of the classifier of interest (daily living score at 14 months) and the other classifiers built on measures at the same time-point (14 months).

Paired classifiers (daily living vs.)	*z*	*p*
msel	-2.19	0.028
vabs	-0.92	0.359
aosi	-2.50	0.013
msel + vabs	-1.17	0.241
msel + aosi	-2.40	0.017
vabs + aosi	-0.30	0.767
all instruments	-2.55	0.011
motor	-2.80	0.005
communication	-0.97	0.333
social	-2.80	0.005
motor + communication	-2.09	0.037
motor + social	-2.40	0.017
communication + social	-1.17	0.241
motor + daily living	-1.07	0.285
communication + daily living	-0.46	0.646
social + daily living	-2.50	0.013
motor + communication + daily living	-2.80	0.005
motor + social + daily living	-2.04	0.041
communication + social + daily living	-1.07	0.285
motor + communication + social	-2.80	0.005
all domains	-2.80	0.005

Notes. Paired tests were performed as post-hoc analysis for significant Friedman tests on predictive performance of classifiers at the same time-point. Bonferroni correction was used to correct for multiple comparisons (21 pairs), and results were significant for $p<\alpha_{Bonferroni}$, with $\alpha_{Bonferroni}==0.0024$. *Abbreviations:* MSEL= Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; AOSI= Autism Observation Scale for Infants.

TABLE S6. ASD classifier performance over time. This table shows results from Friedman test on classifier performance over time (8 months, 8 months + change factor between 8 and 14 months, 14 months) measured by Area Under the Curve (AUC).

Classifier	Friedman Test			Wilcoxon Tests					
				8m + slope vs. 8m		14m vs. 8m		14m vs. 8m + slope	
	χ^2	*df*	*p*	*z*	*p*	*z*	*p*	*z*	*p*
msel	1.90	2	0.39	.	.	.	.	.	.
vabs	10.40	2	0.006	-1.68	0.09	-2.60	0.009*	-2.70	0.007*
aosi	3.80	2	0.15	.	.	.	.	.	.
msel + vabs	9.80	2	0.007	-2.80	0.005*	-2.50	0.013*	-1.89	0.059
msel + aosi	1.80	2	0.41	.	.	.	.	.	.
vabs + aosi	13.40	2	0.001	-2.09	0.036	-2.70	0.007*	-2.80	0.005*
all instruments	0.36	2	0.84	.	.	.	.	.	.
motor	12.67	2	0.002	-0.59	0.55	-2.70	0.007*	-2.81	0.005*
communication	14.60	2	0.001	-2.70	0.007*	-2.80	0.005*	-2.50	0.013*
daily living	7.80	2	0.02	-2.40	0.017*	-2.50	0.013*	-2.09	0.037
social	13.40	2	0.001	-0.76	0.45	-2.70	0.007*	-2.80	0.005*
motor + communication	0.60	2	0.74	.	.	.	.	.	.
motor + social	3.80	2	0.15	.	.	.	.	.	.
communication + social	12.60	2	0.002	-1.07	0.28	-2.70	0.007*	-2.80	0.005*
motor + daily living	4.97	2	0.08	.	.	.	.	.	.
communication + daily living	14.60	2	0.001	-1.48	0.14	-2.80	0.005*	-2.70	0.007*
social + daily living	5.60	2	0.06	.	.	.	.	.	.
motor + communication + daily living	8.60	2	0.014	-0.66	0.51	-2.50	0.013*	-1.87	0.059
motor + social + daily living	7.40	2	0.025	-2.81	0.005*	-1.99	0.047	-1.38	0.17
communication + social + daily living	9.39	2	0.009	-2.40	0.017*	-1.79	0.07	-1.48	0.14
motor + communication + social	3.80	2	0.15	.	.	.	.	.	.
all domains	7.40	2	0.03	-1.78	0.07	-0.76	0.45	-2.60	0.009*

Notes. Post-hoc paired Wilcoxon tests were performed when the effect of time was significant ($p<0.05$ from Friedman test). Results are reported. Bonferroni correction was used to correct for multiple comparisons (3 pairs), and results were significant (') for $p<\alpha_{Bonferroni}$, with $\alpha_{Bonferroni}==0.017$. *Abbreviations:* MSEL= Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; AOSI= Autism Observation Scale for Infants.

PREDICTION OF HR-ASD + HR-ATYPICAL VS. HR-TYPICAL

Table S7 shows performance measured by AUC of different classifiers at predicting *HR-ASD* and *HR-Atypical* vs. typically developing siblings at 8 months, 8 months adding the change factor between 8 and 14 months, and 14 months. We found a significant difference in classifier performance at 8 months, $\chi^2(21)=137$, $p<0.001$; 8 months with the addition of the change factor between 8 and 14 months, $\chi^2(21)=131$, $p<0.001$; and 14 months, $\chi^2(21)=105$, $p<0.001$. Results from post-hoc paired Wilcoxon tests are shown in Table S8. The classifier with highest AUC was built on the integration of Vineland and AOSI scores at 14 month, yet paired Wilcoxon tests between this classifier and the ones with highest AUC at 8 months and 8 months plus the change factor between 8 and 14 months (motor + communication scores) missed significance (respectively $z=-0.7$, $p=0.5$; $z=-0.8$, $p=0.4$). We also tested significant changes on performance of the same classifier over time by means of a nonparametric Friedman test, and Bonferroni corrected post-hoc paired Wilcoxon tests (3 pairs, $\alpha_{Bonferroni}=0.017$). Results are shown in Table S9.

We also compared performance of the classifiers with highest AUC at each time-point between the two different classification problems (prediction of *HR-ASD* and *HR-ASD + HR-Atypical*) by means of paired Wilcoxon tests. Differences missed significance. We compared at 8 months the motor classifier for ASD and the motor + communication classifier for atypical classification ($z=-1.1$, $p=0.3$); at 8 months plus change factor, the motor + social + daily living classifier for ASD classification vs. the motor + communication classifier for atypical classification ($z=-0.8$, $p=0.4$); at 14 months the daily living classifier for ASD classification vs. the VABS + AOSI classifier for atypical classification ($z=-0.3$, $p=0.8$).

TABLE S7. Classifier performances for prediction of ASD plus atypical outcome at 36 months. Predictive performance as measured by the AUC of the classifiers using as features developmental, behavioural and symptoms measures for classifying HR siblings who later develop ASD or with other atypical development from their typically developing (*HR-ASD + HR-Atypical vs. HR-Typical*).

Classifier	8 months		8 months + change factor		14 months	
	p	AUC (%)	*p*	AUC (%)	*p*	AUC (%)
msel	0.02*	66.5 (52.6, 79.8)	0.01*	67.7 (53.8, 80.9)	0.05	64.6 (50.3, 77.9)
vabs	0.047*	63.9 (49.7, 77.3)	0.04*	65.1 (51.2, 78.0)	0.05	65.1 (50.7, 78.4)
aosi	0.24	56.0 (41.8, 69.4)	0.09	62.3 (47.5, 75.7)	0.01*	69.8 (56.5, 82.1)
msel + vabs	0.02*	66.6 (53.0, 79.1)	0.01*	67.8 (53.9, 80.6)	0.03*	66.9 (52.7, 79.8)

TABLE S7 CONTINUED.

Classifier	8 months		8 months + change factor		14 months	
	p	AUC (%)	*p*	AUC (%)	*p*	AUC (%)
msel + aosi	0.02*	67.1 (53.2, 80.2)	0.01*	68.7 (54.9, 81.5)	0.03*	67.6 (53.8, 80.8)
vabs + aosi	0.04*	64.8 (50.6, 78.0)	0.02*	66.8 (52.8, 79.6)	0.01*	70.8 (57.1, 83.2)
all instruments	0.01*	67.6 (53.6, 80.1)	0.01*	68.8 (55.0, 81.5)	0.02*	69.1 (55.1, 82.0)
motor	0.02*	65.9 (51.2, 68.6)	0.02*	66.8 (52.6, 79.5)	0.07	62.6 (48.3, 76.3)
communication	0.03*	65.9 (52.2, 78.9)	0.01*	67.2 (53.1, 79.8)	0.02*	68.1 (53.9, 80.9)
daily living	0.24	56.2 (42.0, 70.2)	0.19	56.7 (41.8, 71.0)	0.07	64.2 (50.3, 77.6)
social	0.42	51.6 (37.0, 66.0)	0.29	54.5 (39.7, 68.6)	0.17	58.5 (43.8, 72.6)
motor + communication	0.01*	69.2 (55.6, 81.8)	0.01*	69.4 (55.6, 82.1)	0.02*	67.5 (53.4, 80.2)
motor + social	0.03*	64.9 (50.9, 78.1)	0.02*	66.2 (52.0, 79.0)	0.07	63.6 (48.8, 76.3)
communication + social	0.03*	65.1 (50.6, 78.4)	0.02*	66.8 (53.1, 79.4)	0.02*	67.8 (54.2, 80.4)
motor + daily living	0.03*	64.3 (50.2, 77.4)	0.03*	64.7 (50.6, 77.7)	0.04*	64.8 (50.7, 77.8)
communication + daily living	0.04*	64.5 (50.5, 77.7)	0.03*	66.0 (51.6, 78.7)	0.02*	68.5 (54.4, 81.5)
social + daily living	0.29	54.9 (40.1, 68.9)	0.24	55.6 (40.8, 69.9)	0.08	63.0 (48.6, 76.4)
motor + communication + daily living	0.01*	68.2 (54.6, 81.0)	0.01*	68.7 (54.9, 81.3)	0.02*	68.1 (53.9, 81.0)
motor + social + daily living	0.047*	63.3 (49.1, 76.5)	0.03*	64.6 (50.4, 77.7)	0.05	63.9 (49.8, 77.5)
communication + social + daily living	0.06	63.6 (49.2, 76.7)	0.04*	65.3 (51.7, 78.3)	0.03*	67.4 (53.6, 80.3)
motor + communication + social	0.01*	68.5 (55.2, 81.3)	0.01*	69.3 (55.5, 82.0)	0.02*	67.2 (52.5, 80.3)
all domains	0.01*	67.7 (53.9, 80.6)	0.01*	68.3 (54.5, 80.8)	0.02*	67.7 (53.7, 80.6)

Notes. The significance of classification AUC was determined by permutation test, the resulting p-values are reported. Prediction was considered different from chance level if $p<0.05$ (marked as *). 95% confidence interval is reported in parentheses. Measures are reported as *mean (lower level CI, upper level CI)*. *Abbreviations:* AUC = area under the curve; MSEL = Mullen Scales of Early Learning (5 scores); VABS = Vineland Adaptive Behavior Scales (4 scores); AOSI = Autism Observation Scale for Infants, in this study we considered the total score.

TABLE S8. Paired Wilcoxon tests on classifier performance for ASD and atypical outcome prediction. This table shows results from paired Wilcoxon tests (*z-score* and two-tailed *p-value*) on predictive performance, measured by Area Under the Curve (AUC), of the classifier of interest (motor and communication scores at 8 months and 8 months plus the change factor between 8 and 14 months; VABS and AOSI scores at 14 months) and the other classifiers built on measures at the same time-point.

Paired classifiers	**motor + communication (8 months)**		**motor + communication (8 months+change factor)**		**VABS + AOSI (14 months)**	
	z	***p***	***z***	***p***	***z***	***p***
msel	-1.99	0.047	-1.68	0.09	-2.80	0.005
vabs	-2.50	0.013	-2.40	0.017	-2.80	0.005
aosi	-2.80	0.005	-2.50	0.013	-0.46	0.65
msel + vabs	-2.29	0.022	-2.35	0.019	-2.70	0.007
msel + aosi	-1.68	0.093	-0.56	0.58	-1.68	0.09
vabs + aosi	-2.29	0.022	-1.68	0.09	.	.
all instruments	-1.78	0.074	-0.97	0.33	-0.97	0.33
motor	-2.70	0.007	-2.70	0.007	-2.80	0.005
communication	-2.09	0.037	-2.29	0.022	-1.89	0.059
social	-2.80	0.005	-2.80	0.005	-2.80	0.005
daily living	-2.80	0.005	-2.80	0.005	-2.60	0.009
motor + communication	.	.	.	.	-2.40	0.017
motor + social	-2.80	0.005	-2.70	0.007	-2.81	0.005
communication + social	-2.60	0.009	-2.70	0.007	-2.29	0.005
motor + daily living	-2.70	0.007	-2.81	0.005	-2.70	0.007
communication + daily living	-2.19	0.028	-2.29	0.022	-1.60	0.11
social + daily living	-2.80	0.005	-2.80	0.005	-2.80	0.005
motor + communication + daily living	-1.48	0.14	-1.84	0.06	-1.99	0.047
motor + social + daily living	-2.80	0.005	-2.80	0.005	-2.80	0.005
communication + social + daily living	-2.35	0.019	-2.40	0.017	-2.70	0.007
motor + communication + social	-2.20	0.028	-1.07	0.28	-2.50	0.013
all domains	-1.99	0.047	-2.30	0.022	-2.50	0.013

Notes. Paired tests were performed as post-hoc analysis for significant Friedman tests on predictive performance of classifiers at the same time-point. Bonferroni correction was used to correct for multiple comparisons (21 pairs), and results were significant for $p<\alpha_{Bonferroni}$, with $\alpha_{Bonferroni}$==0.0024. *Abbreviations.* MSEL= Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; AOSI= Autism Observation Scale for Infants.

TABLE S9. ASD + atypical classifier performance over time. This table shows results from Friedman test on classifier performance over time (8 months, 8 months + change factor between 8 and 14 months, 14 months) measured by Area Under the Curve (AUC).

Classifier	**Friedman Test**			**Wilcoxon Tests**					
				8m + slope vs. 8m		**14m vs. 8m**		**14m vs. 8m + slope**	
	χ^2	***df***	***p***	***z***	***p***	***z***	***p***	***z***	***p***
msel	3.2	2	0.2	.	.	.	.	.	.
vabs	2.0	2	0.4	.	.	.	.	.	.
aosi	12.8	2	0.002	-2.7	0.007*	-2.6	0.009*	-2.5	0.013*
msel + vabs	1.8	2	0.4	.	.	.	.	.	.
msel + aosi	3.8	2	0.2	.	.	.	.	.	.
vabs + aosi	7.8	2	0.02	-2.6	0.009*	-1.9	0.06	-1.5	0.1
all instruments	3.8	2	0.2	.	.	.	.	.	.
motor	7.2	2	0.03	-2.1	0.04	-1.9	0.06	-2.0	0.05
communication	2.4	2	0.3	.	.	.	.	.	.
daily living	8.1	2	0.02	-1.6	0.1	-2,4	0.017*	-2.4	0.017*
social	3.2	2	0.2	.	.	.	.	.	.
motor + communication	0.8	2	0.7	.	.	.	.	.	.
motor + social	2.4	2	0.3	.	.	.	.	.	.
communication + social	4.1	2	0.1	.	.	.	.	.	.
motor + daily living	0.8	2	0.7	.	.	.	.	.	.
communication + daily living	3.2	2	0.2	.	.	.	.	.	.
social + daily living	9.6	2	0.008	-1.1	0.3	-2.3	0.02	-2.8	0.005*
motor + communication + daily living	1.4	2	0.5	.	.	.	.	.	.
motor + social + daily living	2.6	2	0.3	.	.	.	.	.	.
communication + social + daily living	4.2	2	0.1	.	.	.	.	.	.
motor + communication + social	2.6	2	0.3	.	.	.	.	.	.
all domains	1.4	2	0.5	.	.	.	.	.	.

Notes. Post-hoc paired Wilcoxon tests were performed when the effect of time was significant ($p<0.05$ from Friedman test). Results are reported. Bonferroni correction was used to correct for multiple comparisons (3 pairs), and results were significant (*) for $p<\alpha_{Bonferroni}$, with $\alpha_{Bonferroni}$==0.017. *Abbreviations:* MSEL= Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; AOSI= Autism Observation Scale for Infants.

3

Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants

Published as: "Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants", J Autism Dev Disord, 2019.

Pijl, M. K. J.,
Bussu, G.,
Charman, T.,
Johnson, M. H.,
Jones, E. J. H.,
Pasco, G.,
Oosterling, I. J.,
Rommelse, N. N.,
Buitelaar, J. K. & The BASIS Team

ABSTRACT

To investigate temperament as an early risk marker for autism spectrum disorder (ASD), we examined parent-reported temperament for high-risk (HR, n=170) and low-risk (LR, n=77) siblings at 8, 14, and 24 months. Diagnostic assessment was performed at 36 months. Group-based analyses showed linear risk gradients, with more atypical temperament for HR-ASD, followed by HR-Atypical, HR-Typical, and LR siblings. Temperament differed significantly between outcome groups ($0.34 \geq \eta_p^2 \geq 0.03$). Machine learning analyses showed that, at an individual level, HR-ASD siblings could not be identified accurately, whereas HR infants without ASD could. Our results emphasize the discrepancy between group-based and individual-based predictions and suggest that while temperament does not facilitate early identification of ASD individually, it may help identify HR infants who do not develop ASD.

INTRODUCTION

Temperament can be defined as relatively stable individual differences in activity, affectivity, attention, and self-regulation that are shaped throughout development by complex interactions between genetic, biological, and environmental factors [60]. Given that temperament traits can be linked to neurobiological systems [148, 149] and are already measurable at an early age, potentially before psychopathology begins to emerge, temperament could function as a potential risk marker of later psychopathology [61, 150, 151]. The aim of this study was to investigate temperament as an early risk marker for autism spectrum disorders (ASD) in the high-risk (HR) younger siblings of children diagnosed with ASD and low-risk (LR) controls. Research has shown that 18.7% of HR siblings are diagnosed with ASD themselves [27], and that 19% of HR siblings have some traits common to ASD, but not sufficient to warrant a clinical diagnosis [152]. By applying a HR design, shared and unique characteristics of temperament between and within familial HR siblings (diagnosed with ASD, atypically developing, or typically developing) and LR siblings can be studied to reveal possible early predictors of later ASD or atypical development.

Most temperament frameworks encompass three traits during early childhood: (1) *surgency/approach* referring to engagement with the environment, positive emotions, and activity level; (2) *negative affect/withdrawal* including negative emotions such as anger, sadness, and fear; and (3*) effortful control* referring to regulation of attention, emotions, and behaviors [153]. In infancy, effortful control is described as orienting/regulation, focusing on soothability (pace of recovery from distress) and cuddliness (expression of enjoyment and molding of the body to the caregiver) [154]. In the current study, we refer to this construct as *effortful control* in both infancy and toddlerhood.

Previous research has revealed that these three broader traits can differentiate children with ASD from others from 12 months onward (see Table 1). First, low levels of the trait *surgency* (i.e., approach behaviors, positive affect, and activity level) have been associated with later ASD [46, 63, 64, 155, 156]. However, findings up to 1 year are discrepant, showing that HR siblings that develop ASD have *higher* levels of surgency than high-risk siblings who do not develop ASD [64, 65]. This discrepancy suggests that temperamental patterns change with development, but could also reflect differences in the applied construct of surgency across age and as used in different temperament measures. In-depth examination at a dimensional level showed contrasting patterns for activity levels, with lower levels of activity being seen in infants with (or at risk of) ASD during the first year [46, 64, 68], followed by elevated levels of activity around the second year [63, 68]. Second, higher levels of the temperament trait *negative affect* have been consistently associated with ASD from 12

months onward [46, 63, 65, 68, 155, 156]. Lastly, children with ASD have more self-regulatory difficulties (*effortful control*) from around the first birthday onward [46, 63, 65, 66, 68, 155, 156]. However, Del Rosario, Gillespie-Lynch [64] did not find any differences in negative affect or effortful control between HR-ASD and LR siblings during early childhood, which could be due to the use of different instruments to assess temperament. See Table 1 for a detailed overview of the abovementioned studies focusing on temperament traits in ASD.

TABLE 1. Summary of findings on the three temperament traits and/or dimensions related to the traits in infants and toddlers with (or at risk of) ASD.

Study	Participant description (N)	0-11 months			1-2 years			2-3 years		
		SU	NA	EC	SU	NA	EC	SU	NA	EC
Clifford et al., 2013	HR-ASD (17), HR-Atypical (12), HR-Typical (24), LR (50)	↑[1]	ns	ns	↑[1]↓[2]	ns	↓[3]	ns	↑[2]	↓[2]
Del Rosario et al., 2014	HR-ASD (10-16), HR-non ASD (7-27)	↑↓	ns	ns	↓	ns	ns	↓	ns	ns
Garon et al., 2009	HR-ASD (34), HR-non ASD (104), LR (73)							↓↑[4]	↑[2]	↓[4]
Garon et al., 2016	HR-ASD (98), HR-non ASD (285), LR (162)				↓[5]	↑[6]	ns	↓[5,6]	↑[6]	↓[5,6]
Zwaigenbaum et al., 2005	HR-ASD (19), HR non-ASD (46), LR (23)	↓[4]	ns	ns	ns	↑[4]	↓[4]	↓[4]	ns	↓[4]
Gomez & Baird, 2005	ASD (65), TD (120)						↓			
Bolton et al., 2012	ASD (85), non-ASD (13885)	↓[7]	ns	ns				↑[7]	↑[7]	↓[7]
Macari et al., 2017	ASD (165), DD (58), TD (92)							↓[8]	↑[9]	↓[8]

Notes. ADOS = Autism Diagnostic Observation Schedule; ADI-R = Autism Diagnostic Interview - Revised; ASD = infants or toddlers diagnosed with autism spectrum disorders; DD = developmentally delayed infants or toddlers; DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders, fourth version text revision; EC = Effortful Control; HR-ASD = at-risk siblings subsequently diagnosed with ASD; HR-Atypical = at-risk siblings not diagnosed with ASD, but following an atypical development; HR-Typical = at-risk siblings following a typical development; ICD-10 = International Classification of Diseases, tenth version; LR = low-risk controls; MSEL = Mullen Scales of Early Learning; NA = Negative Affect; SU = Surgency; TD = typically developing infants or toddlers; Vineland = Vineland Adaptive Behavior Scales; Marked cells indicate findings based on the temperament trait's composite score instead of findings based on dimensions or constructs related to the broader trait. Dimensions or constructs that could not be related to one of the three traits were not included in this table; Empty cells not investigated. [1] HR-ASD as compared to HR-Typical; [2] HR-ASD as compared to LR; [3] HR-ASD as compared to LR and HR-Atypical; [4] HR-ASD as compared to HR non-ASD and LR; [5] HR-ASD as compared to HR non-ASD; [6] HR (HR-ASD and HR non-ASD) as compared to LR; [7] Findings reported here are controlled for gender; [8] ASD as compared to both DD and TD; [9] ASD as compared to TD.

Most of the abovementioned studies focused on differences in distinct temperament traits at separate time points (e.g., the level of surgency at 12 months) and did not integrate findings across various traits and time. To the best of our knowledge, only two studies investigated the time course of temperament in young children at risk of ASD [64, 155]. The investigation of trajectories of temperament across multiple time points is potentially more informative than measures of temperament at single time points, because it provides information about the structure of change across early childhood. In addition, investigating the integration of different temperament traits at different time points could help to combine complementary information across traits. Furthermore, while previous studies investigated temperamental differences between groups, they did not examine whether temperament provides information about individual outcomes. Although findings on group differences are valuable in terms of finding relevant biomarkers for ASD, there is often substantial overlap between groups in individual variation, making prediction for individual infants difficult. To fully judge whether temperament is useful in the early prediction of ASD, analyses at an individual level are needed.

The current study prospectively followed familial HR and LR siblings during their first 3 years of life, with the aim of observing differences in temperament between outcome groups. For these outcome groups, the HR group was divided into HR-ASD (HR siblings subsequently diagnosed with ASD at 36 months), HR-Atypical (HR siblings not diagnosed with ASD, but with some evidence of atypical development) and HR-Typical siblings (HR siblings with typical development). The objectives were 1) to investigate *group* differences in early temperament at and across multiple time points between HR-ASD, HR-Atypical, HR-Typical, and LR siblings, and 2) to examine whether temperament (both single traits and profiles) during the first 2 years of life (both separate time points and trajectories) can help to predict ASD and atypical development at 36 months at an *individual* level. For the latter objective we extended previous work by using machine learning algorithms to combine complementary information about different temperament factors in order to identify the best predictive combination of factors. We expected that trajectories of temperament would differentiate between outcome groups and that the integration of different domains of temperament measured at different time points would improve the prediction of ASD in an individual as compared to prediction based on a single domain and/or time point. Further, based on their risk status, we hypothesized that HR-ASD would show the most 'atypical' temperament (i.e., low surgency, high negative affect, low effortful control), followed by HR-Atypical siblings, HR-Typical siblings, and LR siblings.

METHODS

PARTICIPANTS AND PROCEDURE

As part of the British Autism Study on Infant Siblings (BASIS: www.basisnetwork.org), 247 infants (170 HR and 77 LR) were assessed at four time points during their first three years of life. Data for 104 infants were collected during the first phase of the longitudinal study, which were also reported by Clifford, Hudry [65]. Ethical approval was given by NHS NRES London RC (06/MRE02/73, 08/H0718/76), and one or both parents gave informed consent. Most of the infants were born full-term (i.e., N=236 were born ≥36 weeks, N=11 were born between 32 and 36 weeks) and none of them had known medical or developmental conditions at the time they were enrolled. The HR infants had at least one older sibling with a clinical diagnosis of ASD (hereafter: 'proband'), confirmed in most cases by expert clinicians using information from the Development and Wellbeing Assessment [DAWBA - 157] and the Social Communication Questionnaire [SCQ - [158]]. No known other significant conditions were present in the proband or extended family members (e.g., Fragile X syndrome, tuberous sclerosis). LR siblings were recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. There was no ASD in first-degree family members of LR siblings (as confirmed through a parent interview regarding family medical history).

Of 247 siblings recruited, data for 33 HR and 9 LR siblings were excluded from the current study because of a substantial amount of missing data. Further information about this exclusion criterion is presented in the Measures section. We also excluded infants with no information about outcome status (N=4 HR, N=2 LR). The final sample comprised 133 HR infants (65 male; 48.9%) and 66 LR infants (28 male; 42.4%). All infants were examined at approximately 8 months (mean=8.4, SD=1.3, hereafter 8 months), 14 months (mean=14.8, SD=1.4, hereafter 14 months), around their second birthday (mean=25.4, SD=1.9, hereafter 24 months), and around their third birthday (mean=38.6, SD=2.2, hereafter 36 months).

MEASURES

Infant and toddler temperament

Two measures of temperament, appropriate to the child's age, were administered. Parents completed the Infant Behavior Questionnaire-Revised [IBQ-R - [154]] at the 8- and 14-month visits, and the Early Childhood Behavior Questionnaire [ECBQ - [159]] at the 24-month visit. Both measures are reliable and well-validated parent-reported questionnaires that are scored on a Likert scale ranging from 1 (never) to 7 (always). The IBQ-R was designed to assess temperament in the first year of life and contains 14 dimensions based on 184 items. The ECBQ was developed for children aged 18 to 36 months and consists of 18 dimensions based on 201 items. Three broad factors can be identified with both the IBQ-R and the ECBQ: Surgency, Negative Affect, and Effortful Control (labeled 'Orienting' on the IBQ-R). Of note, although both the IBQ-R and ECBQ provide a similar 3-factor model, the loading on the factors is different. See Putnam, Ellis [153] for a discussion of this structure of temperament.

To ensure the validity of the temperament measures, dimensions were only calculated if data on ≥70% of items were available. Similarly, factors were only computed if ≥70% of dimension scores were available. Given that this study focused on longitudinal trajectories of temperament at 8, 14, and 24 months, participants were only included if data on ≥70% of the factors were available across the three time points.

Outcome characterization

At the 36-month visit, various clinical research measures were used to characterize the outcome of the HR siblings. The Autism Diagnostic Observation Schedule [ADOS-2 - [160]], the Autism Diagnostic Interview [ADI-R - [161]], and the SCQ [158] were used to obtain information about ASD symptomatology. In addition, the Mullen Scales of Early Learning [1] and the Vineland Adaptive Behavior Scale-II [2] were assessed to gather information about the child's development and adaptive functioning, respectively. Experienced clinical researchers (TC, GP) reviewed the outcomes of each HR sibling. Consensus ICD-10 or DSM-5 criteria were used to ascertain ASD diagnostic outcome. Among the 133 HR siblings enrolled in this study, 24 HR siblings met criteria for ASD (hereafter: 'HR-ASD') and 34 HR siblings did not meet criteria for ASD, but scored above the ASD threshold on the ADOS and/or ADI-R and/or scored >1.5 SD below the population mean on the MSEL receptive language, expressive language, and/or early learning composite score [hereafter: 'HR-Atypical']). The remaining 75 HR siblings were considered to be developing typically (hereafter: 'HR-Typical'). No formal research diagnoses were assigned to the LR group, but none of the LR infants had a community clinical ASD diagnosis. See Table 2 for detailed demographics of the included participants.

TABLE 2. Sample characterization (means and standard deviations) for low-risk siblings and subgroups of high-risk siblings.

	HR-ASD (N=24)	HR-Atypical (N=34)	HR-Typical (N=75)	LR (N=66)
Sex (% male)	75[a]	47.1	41.3[b]	42.4[b]
Age				
8 months	8.3 (1.4)	8.6 (1.0)	8.5 (1.3)	8.3 (1.4)
14 months	14.8 (1.6)	14.7 (1.4)	14.9 (1.3)	14.7 (1.3)
24 months	25.4 (2.8)	25.4 (2.1)	26.0 (1.9)[a]	24.7 (1.0)[b]
36 months	38.0 (2.0)	38.0 (2.8)	38.5 (1.8)	38.4 (2.7)
MSEL[1]				
8 months	98.0 (15.5)[a]	100.0 (13.8)	106.3 (15.8)	107.7 (12.6)[b]
14 months	89.8 (17.3)[a]	96.5 (14.0)[a]	99.8 (14.6)	106.0 (15.0)[b]
24 months	94.5 (24.8)[a]	99.2 (21.8)[a]	104.9 (15.9)[a]	115.4 (14.2)[b]
36 months	98.0 (26.7)[a]	95.9 (24.4)[a]	115.1 (15.5)[b]	118.1 (15.0)[b]
ADOS severity[2,3]				
36 months	5.1 (3.0)[a]	5.1 (2.2)[a]	1.5 (0.9)[b]	2.5 (1.8)[c]

Superscripted letters that differ from other superscripted letters indicate significant differences across groups for the given measure ($p \leq 0.05$). Values without superscript letters indicate no significant differences from another group.

[1] Mullen Scales of Early Learning (Mullen, 1995) Early Learning Composite Standard Score

[2] Autism Diagnostic Observation Schedule-2 (ADOS-2 - Lord et al., 2012)

[3] ADOS-2 calibrated severity score (Gotham et al., 2009)

STATISTICAL ANALYSES

Multiple imputation with the expectation maximization algorithm was used to account for missing data [162]. In addition, a Van der Waerden transformation was applied to data for temperament factors, which transforms raw scores into z-scores corresponding to the estimated cumulative proportion of the distribution analogous to a particular rank (using Statistical Package for the Social Sciences [SPSS] version 22).

Group-based analyses

MANCOVAs were used to investigate whether a risk gradient was present in polynomial group contrasts at separate time points. The outcome groups were ranked as follows: 1=HR-ASD, 2=HR-Atypical, 3=HR-Typical, and 4=LR, assuming that polynomial group contrasts would indicate linear risk gradients for atypical temperament (HR-ASD > HR- Atypical > HR-Typical > LR). Analyses were performed for each temperament trait separately, including group as independent variable and temperament at three time points as dependent variables (e.g., surgency at 8, 14, and 24 months). Sex was differently distributed across groups (with more males than females in the HR-ASD group), and age at intake was variable (between 5 and 11 months), introducing potential noise in results due to different starting ages. Therefore, sex and age at the first visit were included as covariates.

In post-hoc analyses, pair wise group contrasts were examined across time by performing two-way mixed ANCOVAs and paired sample t-tests, resulting in six pair wise comparisons (i.e., HR-ASD vs. HR-Atypical, HR-ASD vs. HR-Typical, HR-ASD vs. LR, HR-Atypical vs. HR-Typical, HR-Atypical vs. LR, HR-Typical vs. LR). The effect of group (e.g., HR-ASD, LR), time (8, 14, 24 months), and the interaction effect group x time on trajectories of a temperament trait was investigated, while controlling for sex and age at the first visit. A correction for multiple comparisons was applied for the post-hoc analyses, using the false discovery rate controlling procedure with a q-value of 0.05 [163]. If Mauchly's test indicated that the assumption of sphericity had been violated, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Following Cohen's guidelines (Cohen, 1988), effect sizes were defined in terms of the percentage of variance explained: 1, 9 and 25% were used to define small, medium, and large effects (these percentages translate into ηp2-values of 0.01, 0.06 and 0.14). Analyses contrasting the HR group (without a differentiation based on 36-month outcome) and the LR controls are described in *Supplemental Material*.

Classifier Analyses – from group-based to individual analysis

As a next step, we investigated how temperament factors at 8, 14, and 24 months related to atypical development, and more specifically ASD, at an individual level among infants in the HR group. To this end, we performed confounder-corrected support vector machine classification with 40% holdout cross-validation repeated 10 times using custom made scripts implemented in Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016). We addressed two distinct binary classification problems: distinguishing HR-ASD from HR-Atypical and HR-Typical; and distinguishing HR-ASD and HR-Atypical together from HR-Typical. In fact, while the most clinically relevant question is to distinguish HR-ASD siblings from their peers at an early age, distinguishing HR-ASD and HR-Atypical together from HR-Typical is also clinically relevant and potentially useful for early intervention. Sex and age at the first visit were included as covariates, and findings were corrected for inverse probability weighting. Features for the classifiers consisted of temperament factors (surgency, negative affect, effortful control, and all their combinations) from different time points (8 months, 14 months, and 24 months). To exploit the longitudinal information on developmental dynamics, the intercept and slope of the developmental trajectories on single measures between 8 and 24 months were also used as features for the classifiers. Trajectories were computed for single individuals by linear regression modeling using the

lme4 software package on R [140]. A total of 28 classifiers were compared to find the best predictor of HR-ASD and HR-ASD+HR-Atypical at 36 months (see *Supplemental Material* for details). For each classifier, the area under the curve (AUC) was computed to determine the best classifier, and we evaluated the classifier performance via sensitivity, specificity, accuracy, negative predictive value (NPV – i.e., true negative over negative predicted cases), and positive predictive value (PPV – i.e., true positive over positive predicted cases). 95% confidence intervals (CI) for each metric were computed using bootstrap with n=1000 repetitions for each cross-validation fold, then averaging over folds (n=10000 in total). The *p*-value of AUC was computed for each classifier through a shuffle test (n=10000 total repetitions; n=1000 repetitions for each classification fold) to test the significance of classification performance. Performance metrics are reported only when the performance was significantly different from chance level.

For both classifications (HR-ASD vs. HR-Atypical + HR-Typical | HR-ASD + HR-Atypical vs. HR-Typical), the best predicting classifier at each time point was selected based on the AUC. A nonparametric Friedman test was performed on classifier performance metrics (i.e., AUC) at each time point separately to test for significant differences in performance between distinct classifiers. If the Friedman test was significant, post-hoc paired Wilcoxon rank sum tests were performed between the classifier of interest (i.e., the one with highest AUC) and all other classifiers. Bonferroni correction was applied to avoid biasing effects due to multiple comparisons. In addition, differences in performance of the best classifiers across time points were tested by a two-sided Wilcoxon rank sum test.

RESULTS

TEMPERAMENT DIFFERENCES BETWEEN GROUPS

Surgency

A polynomial group contrast indicated a linear risk gradient to be present at 14 months of age (Contrast Estimate [CE]=0.40, p=0.02), implying that LR siblings had the highest levels of surgency, followed by HR-Typical siblings, HR-Atypical siblings, and HR-ASD siblings. No significant gradient was present at 8 or 24 months (CE=-0.08, p=0.64; CE=0.27, p=0.10, respectively).

Two-way mixed ANCOVAs examining pair wise group contrasts revealed a significant group x time effect for the comparison between HR-ASD and HR-Typical siblings (F(1.77, 168.07)=3.67, p<0.05, η_p^2=0.04; see Figure 1), as well as between HR-ASD and LR siblings (F(1.86, 160.06)=3.98, p<0.05, η_p^2=0.04). Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age (F(1, 95)=6.69, p<0.05, η_p^2=0.07; F(1, 86)=9.79, p<0.01, η_p^2=0.10, respectively), with HR-ASD siblings showing diverging levels of surgency (i.e. approach behaviors, positive affect, activity level) from 8 to 14 months compared with HR-Typical and LR siblings (paired sample t-tests for each group were non-significant). In addition, for the comparison between HR-ASD and LR siblings a significant main effect of group was found between the 14- and 24-month time point (F(1, 86)=4.89, p<0.05, η_p^2=0.05), indicating stable lower levels of surgency in the HR-ASD group than in the LR group between 14 and 24 months of age.

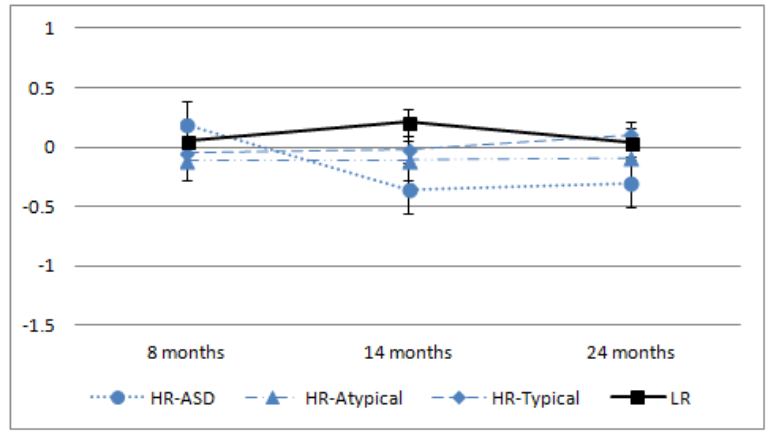


FIGURE 1. Estimated Means for Surgency by Diagnostic Group and Time controlled for Sex and Age at Start. This figure shows the estimated means for surgency with error bars representing standard errors. A significant group x time effect was found for the comparison between HR-ASD

and HR-Typical siblings, and between HR-ASD and LR siblings. Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age. Additionally, for the comparison between HR-ASD and LR siblings a significant main effect of group was found between the 14- and 24-month time point. For details see text.

Negative Affect

A polynomial group contrast indicated a linear risk gradient to be present at 8, 14, and 24 months (CE=-0.46, p=0.004; CE=-0.38, p=0.02; CE=-0.69, p<0.001, respectively), suggesting that HR-ASD siblings showed the highest levels of negative affect, followed by HR-Atypical siblings, HR-Typical siblings and LR siblings.

A two-way mixed ANCOVA revealed significant main group effects for HR-ASD vs. HR-Typical (F(1, 95)=7.47, p<0.01, η_p^2=0.07; see Figure 2), HR-ASD vs. LR (F(1, 86)=15.57, p<0.001, η_p^2=0.15), and HR-Typical vs. LR siblings (F(1, 137)=6.49, p<0.05, η_p^2=0.05). These effects indicate that, independent of age, HR-ASD siblings had developmentally stable higher levels of negative affect than HR-Typical and LR siblings, and that HR-Typical siblings had stable higher levels of negative affect than LR siblings.

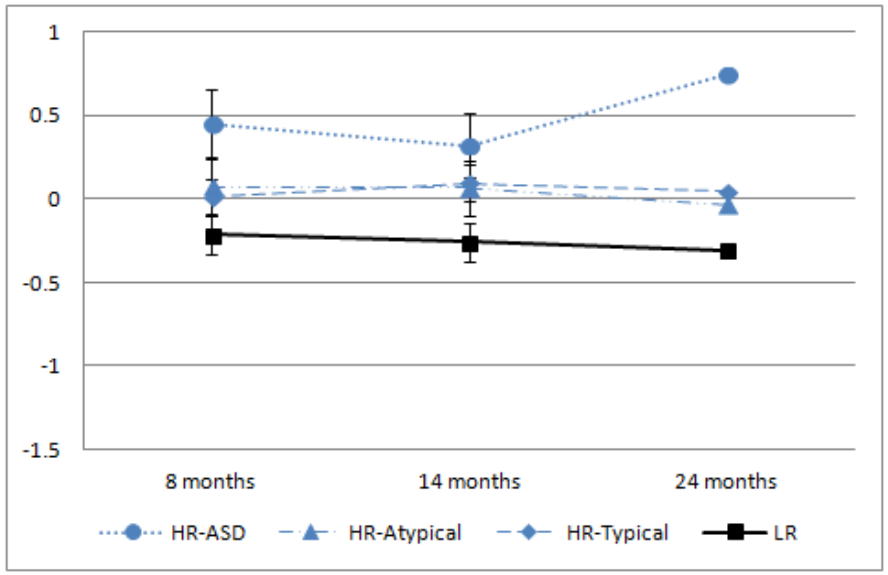


FIGURE 2. Estimated Means for Negative Affect by Diagnostic Group and Time controlled for Sex and Age at Start. This figure shows the estimated means for negative affect with error bars representing standard errors. Significant main group effects were found for HR-ASD vs HR-Typical, HR-ASD vs LR, and HR-Typical vs LR siblings. For details see text.

Effortful Control

A polynomial group contrast indicated a linear risk gradient to be present at 14 and 24 months (CE=0.69, p<0.001; CE=0.84, p<0.001, respectively), showing that LR siblings had the highest levels of effortful control, followed by HR-Typical siblings, HR-Atypical siblings and HR-ASD siblings. No significant gradient was present at 8 months of age (CE=0.21, p=0.20).

A two-way mixed ANCOVA showed significant group x time interaction effects for the comparisons between HR-ASD and HR-Typical siblings (F(1.85, 175.79)=6.95, p<0.01, η_p^2=0.07; see Figure 3), and between HR-ASD and LR siblings (F(2, 172)=8.41, p<0.001, η_p^2=0.09). Post-hoc tests revealed that the interaction effects were driven by the 8- to 14-month trajectory (F(1, 95)=8.53, p<0.01, η_p^2=0.08; F(1, 86)=12.69, p<0.01, η_p^2=0.13, respectively), showing that the level of effortful control decreased in HR-ASD siblings from 8 to 14 months (t(23)=2.85, p=0.009) relative to the static levels of effortful control seen in HR-Typical (t(74)=-1.08, p=0.28) and LR (t(65)=-1.03, p=0.31) siblings. Between 14 and 24 months of age, significant main effects of group were found (F(1, 86)=18.90, p<0.001, η_p^2=0.17; F(1, 86)=44.22, p<0.001, η_p^2=0.34, respectively), suggesting that HR-ASD siblings had stable lower levels of effortful control than HR-Typical and LR siblings. Furthermore, significant main group effects were found between HR-ASD vs. HR-Atypical (F(1, 54)=6.28, p<0.05, η_p^2=0.10), HR-Typical vs. LR (F(1, 137)=4.31, p<0.05, η_p^2=0.03), and HR-Atypical vs. LR (F(1, 96)=5.19, p<0.05, η_p^2=0.05) siblings. These results showed that HR-ASD siblings had developmentally stable lower levels of effortful control than HR-Atypical siblings, and that LR controls had higher levels of effortful control than both HR-Typical and HR-Atypical siblings.

INDIVIDUAL PREDICTION OF HR CLINICAL OUTCOME

Classification of HR-ASD among HR siblings was significantly different from chance level using measures from 14 months onward. In contrast, classification of HR-ASD and HR-Atypical together from HR-Typical was not significantly different from chance level at any of the time points, with only marginal significance at 24 months. See Table 3 and 4 for an overview of the performance metrics of classifiers that were significantly different from chance level for the two classifications (i.e., HR-ASD vs. HR-Atypical + HR-Typical, and HR-ASD+HR-Atypical vs. HR-Typical). Detailed statistics can be found in the *Supplemental Material*.

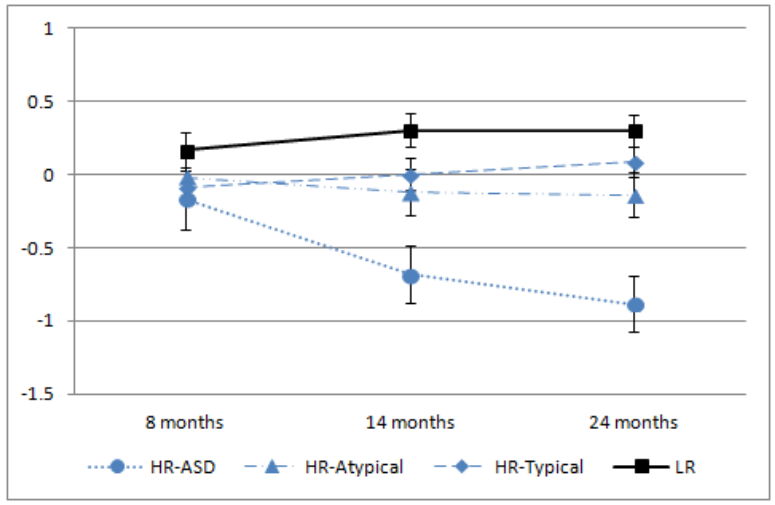


FIGURE 3. Estimated Means for Effortful Control by Diagnostic Group and Time controlled for Sex and Age at Start. This figure shows the estimated means for effortful control with error bars representing standard errors. Significant group x time effects were found for the comparison between HR-ASD and HR-Typical siblings, and between HR-ASD and LR siblings. Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age. Significant main effects of group were found between the 14- and 24-month time point for both comparisons. Additionally, significant main effects of group were found for HR-ASD vs HR-Atypical, HR-Typical vs LR, and HR-Atypical vs LR siblings. For details see text.

TABLE 3. HR-ASD vs. HR-Typical + HR-Atypical. Performance metrics of temperament factors for classifying HR siblings who later develop ASD from their peers.

	Classifier	*p*	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
14 months	Effortful control	0.047	64.4 [46.7, 80.3]	69.6 [37.4, 97.8]	59.2 [44.5, 73.4]	60.9 [47.9, 73.6]	26.6 [9.9, 45.2]	90.3 [78.3, 99.4]
24 months	Effortful control	0.006	71.4 [57.0, 82.6]	88.0 [66.2, 100]	54.8 [39.8, 69.0]	60.6 [47.0, 73.0]	29.2 [12.7, 46.5]	95.6 [87.6, 100]
	Surgency + Effortful control	0.021	66.7 [49.6, 80.8]	79.3 [49.3, 100]	54.1 [39.1, 68.8]	58.5 [44.9, 71.7]	26.7 [10.7, 44.3]	92.6 [80.8, 100]
	Effortful control + Negative affect	0.031	69.9 [55.0, 82.6]	82.6 [57.0, 98.9]	57.3 [42.4, 71.8]	61.7 [48.5, 74.7]	28.9 [12.3, 47.2]	94.0 [84.5, 99.7]
	All factors	0.020	71.5 [57.1, 83.5]	84.8 [60.5, 98.9]	58.2 [43.4, 72.6]	62.8 [49.4, 75.5]	29.9 [13.0, 48.5]	94.8 [85.7, 99.7]
Longitudinal trajectory	Effortful control	0.042	66.4 [49.2, 80.3]	79.3 [48.9, 100]	53.5 [38.9, 67.9]	57.9 [44.7, 70.6]	26.9 [11.2, 44.0]	92.3 [80.2, 100]
	Effortful control + Negative affect	0.013	67.6 [50.5, 81.7]	77.3 [47.3, 99.1]	57.8 [42.8, 72.1]	61.1 [47.9, 74.0]	28.0 [11.5, 46.2]	92.4 [81.0, 99.7]
	All factors	0.048	65.1 [48.1, 79.3]	75.2 [44.8, 97.9]	55.0 [40.2, 69.5]	58.5 [44.9, 71.7]	26.0 [9.9, 43.8]	91.5 [79.3, 99.3]

Notes. All classifiers reported in this table significantly differed from prediction at chance level (shuffle test $p<0.05$). All metrics are reported as *mean [95% confidence interval]*. 95% confidence interval was computed using bootstrap. The classifiers are divided based on the data used as features: data collected at 14 months, data collected at 24 months, intercept and slope of the longitudinal trajectory between 8 and 24 months at the individual level. The abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

TABLE 4. HR-ASD + HR-Atypical vs. HR-Typical. Performance metrics of temperament factors for classifying the HR atypical group as a whole (including atypically developing siblings and those who later develop ASD) from typically developing siblings.

	Classifier	*p*	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
24 months	Surgency + Effortful control	0.058	60.6 [47.0, 73.5]	63.7 [43.0, 82.9]	57.6 [39.6, 74.7]	60.2 [46.8, 72.8]	52.7 [34.2, 70.9]	68.2 [49.1, 85.4]
	Effortful control + Negative affect	0.056	61.1 [48.0, 74.1]	59.8 [39.2, 79.2]	62.4 [45.4, 78.9]	61.3 [48.3, 74.3]	55.0 [34.2, 70.9]	67.9 [49.9, 84.7]
	All factors	0.051	60.0 [46.3, 72.9]	58.8 [37.7, 78.7]	61.2 [43.7, 77.6]	60.2 [47.0, 72.6]	53.2 [33.0, 72.0]	66.7 [48.5, 83.7]

Notes. None of the classifiers performed significantly different from chance level (shuffle test $p<0.05$). Here we report classifiers performing marginally different from random. All metrics are reported as *mean [95% confidence interval]*. 95% confidence interval was computed using bootstrap. Abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

To evaluate which combination of temperament factors best predicted ASD at different time points, we compared the performance of the different classifiers at the separate time points, based on the AUC. The combination of all factors at 24 months provided the most promising classifier to predict ASD among HR siblings (p=0.02; mean [CI]: AUC=72% [57% to 83%]; sensitivity=85% [61% to 99%], specificity=58% [43% to 73%], accuracy=63% [49% to 75%], PPV=30% [13% to 49%], NPV=95% [86% to 100%]). However, the predictive performance was not significantly different from that of effortful control (z=-0.51, p=0.61) and its combination with other factors at 24 months (surgency + effortful control: z=-1.58, p=0.11; effortful control + negative affect: z=-0.98, p=0.33). Furthermore, effortful control had the highest predictive power at 14 months (AUC=64%), and when using the developmental trajectory between 8 and 24 months as feature for the classifiers, the integration of scores from effortful control and negative affect provided the classifier with the highest AUC (AUC=68%). After Bonferroni correction for multiple comparisons (leading to $\alpha_{Bonferroni}$ =0.017), the difference in classification performance between the combined factors at 24 months and effortful control at 14 months was not significant (Wilcoxon z=-2.14, p=0.032), and the same applies to the difference in classification performance between the combined factors at 24 months and the combined longitudinal trajectories of effortful control and negative affect (Wilcoxon z=-1.86, p=0.063).

For classification of HR-ASD plus HR-Atypical from HR-Typical, the integration of effortful control and negative affect at 24 months provided the highest AUC (p=0.056; mean [CI]: AUC=61% [48% to 74%]; sensitivity=60% [39% to 79%], specificity=62% [45% to 79%], accuracy=61% [48% to 74%], PPV=55% [35% to 75%], NPV=68% [50% to 85%]). Since performance was not significantly different from chance level, classifier comparison was not performed.

Overall, even though effortful control and a combination of the temperament factors at 24 months predicted ASD outcome at a moderate level (AUC=71%; AUC=72%, respectively), its positive predictive value for ASD was low and none of the classifiers adequately predicted broader atypical development at 36 months.

DISCUSSION

The current study is the first to examine differences in temperament at and across three time points in early childhood between outcome groups (i.e. HR-ASD, HR-Atypical, HR-Typical and LR siblings), and to investigate temperament at an individual level. At a group level, our findings revealed positive linear risk gradients for surgency at 14 months, and effortful control at 14 and 24 months, and negative linear risk gradients for negative affect at 8, 14, and 24 months, implying that temperament in early childhood was more atypical in HR-ASD siblings, followed by HR-Atypical siblings, HR-Typical siblings, and LR controls. Post-hoc pair wise comparisons indicated differences in early temperament between the outcome groups. However, the effect sizes were generally small, especially regarding differences within the HR group. Machine learning analyses using temperament traits during infancy (i.e., 8 months) did not accurately predict ASD at 36 months at an individual level. From 14 months onward, effortful control (or its combination with other traits) had the highest predictive power for ASD as compared to other temperament traits and combinations, with a high negative predictive value, but with a positive predictive value that was far from being clinically useful. Neither the separate temperament traits nor a combination of traits was able to accurately predict broader atypical development (i.e., HR-ASD and HR-Atypical). Thus, although differences in temperament traits can be detected in infancy at a group level, this difference does not necessarily translate into an acceptably accurate prediction of ASD in the individual infant.

TEMPERAMENT DIFFERENCES BETWEEN HR SUBGROUPS AND LR CONTROLS

At a group level, our findings showed that HR siblings with or without a subsequent ASD diagnosis could be distinguished from LR controls based on higher levels of negative affect and lower levels of effortful control (with the exception of HR-Atypical siblings regarding negative affect). These findings replicate and extend previous research [155], showing that young siblings at risk of ASD, regardless of whether they develop ASD or not, tend to use more negative emotions and have more difficulties regulating attention, emotions, and behaviors than do LR controls. Furthermore, we found that the pattern of surgency from 8 to 14 months and levels of surgency thereafter were different between HR-ASD and LR siblings, whereas levels of surgency in the HR-Typical and HR-Atypical siblings did not differ from those of the LR group. As to be expected, this suggests that, on average, low levels of approach and positive emotions are specifically associated with the development of ASD. Differences in surgency levels across time may be explained by the multi-dimensional nature of the factor surgency [154, 159]. Future research may use a dimensional or item level approach to delineate the underlying mechanisms and to enable comparison of findings between studies.

TEMPERAMENT DIFFERENCES WITHIN AT-RISK SIBLINGS

Within the HR group, temperament traits distinguished HR-ASD siblings from HR siblings without a clinical diagnosis, suggesting the presence of more temperamental challenges early in life of children with subsequent ASD. Interestingly, higher levels of negative affect were already present from 8 months onward in the HR-ASD siblings, whereas effortful control started to distinguish between the groups from 14 months onward. These findings, combined with those of a recent study examining temperament trajectories from 12 months onward [155], may indicate that early affective behaviors play an important role in the subsequent regulation of attention, emotions, and behaviors. Garon, Zwaigenbaum [155] found that affective components of temperament at 12 months predicted regulatory behaviors at 24 months in both HR and LR infants, and that regulatory behaviors in turn predicted ASD symptoms at 36 months in the HR sample. Future investigation of the associations between temperament traits in different outcome groups is needed, including the assessment of temperament during the first year of life.

TEMPERAMENT AS A POTENTIAL EARLY RISK MARKER

The idea that temperament may be an early risk marker is in accordance with the spectrum theory [164], which holds that there is a shared etiology between psychopathology at the extreme negative end of a continuum of social-communicative competences and temperament traits. A study of monozygotic and dizygotic adult twins supported this idea by showing that ASD and most temperament traits share common genetic and environmental etiological factors [165]. Temperament may be a fruitful risk marker that could help differentiate between groups of children on different developmental pathways.

Nonetheless, the use of temperament traits as an early risk marker is constrained by two findings. First, identification of ASD at an individual level on the basis of temperament traits had low positive predictive value and specificity. This indicates that based on (combinations of) temperament traits a substantial number of HR siblings would be falsely classified as HR-ASD at 36 months (i.e., false positives). However, the high negative predictive values indicate that temperament traits can accurately predict which infants are *not* going to develop ASD in all likelihood. This has still clinical value, especially for the selection of infants who might need early intervention. In other words, results at the individual level suggest that while low levels of effortful control do not predict ASD development, high levels of effortful control accurately predict typical development. The predictive value of effortful control for non-ASD development is in line with the view that effortful control, as a measure of executive function, might promote resilience, such that infants with higher levels of effortful control may be better able to compensate for atypicalities that lead to ASD outcome [166]. However, our results highlight the difficulties of translating findings

from a group to an individual level. In fact, there is often substantial overlap between groups in individual variation, making it more difficult to make predictions for individual infants. Instead of a risk marker for ASD, variation in temperament may therefore function as a *stratification marker* that allows to classify individuals with ASD into biologically more homogeneous subtypes [167]. In this way, temperament may help to unravel the heterogeneous character of ASD. Importantly, the extent to which atypical temperament reflects brain alterations that predispose a child to developing ASD and/or are shared between atypical temperament and ASD need to be investigated. Additionally, future work should investigate the integration of clinical (e.g., MSEL, VABS, AOSI) and biological (e.g., eye tracking, functional imaging) measures, to improve the positive predictive value for the clinical diagnosis of ASD at an individual level [55], and to investigate the additional value of temperament. Second, the differences found in this study mainly started to emerge around the first birthday (at both group and individual levels), which is also when behaviors related to ASD start to emerge [135, 168]. This makes it important to ascertain whether temperament measures actually assess characteristics of temperament, or whether they just pick up the emergence of ASD symptoms. Future research should further investigate the conceptual nature of temperament measures by examining the structure of traits in different outcome groups and in relation to ASD severity.

LIMITATIONS AND FUTURE DIRECTIONS

Particular strengths of this study are its longitudinal design, which allowed the assessment of temperament trajectories across early childhood, and the differentiation between siblings based on their diagnostic status at 36 months of age. A limitation is that temperament was assessed on the basis of parent-reported measures and not on observational measures of temperament [e.g., Lab-TAB; 169]. It will therefore be essential to demonstrate convergence between the parent-reported IBQ-R and ECBQ and indicators of temperament based on standardized laboratory or home assessments. Nonetheless, evidence of convergent validity between a preliminary version of the IBQ and home observations of infant temperament implies that parents' familiarity with a child's behavior may make them the best possible source of reliable information [170]. In addition, given that temperament is the result of complex interactions between genetic, biological, and environmental factors (Goldsmith et al., 2006; Shiner et al., 2012), the role of the environment, such as the child's family, should also be considered in temperament research. Previous research has shown that the quality of parenting interacts with individual differences in genetic variation to influence temperament traits [171, 172].

CONCLUSIONS

Taken together, our longitudinal study identified differences in early temperament traits

between HR and LR siblings as well as between the different outcome subgroups among HR children, as most clearly demonstrated by differences in negative affect from 8 months onward and effortful control from 14 months onward. Our results underscore the complexity of translating findings from a group to an individual level, as findings did not accurately predict ASD at an individual level. From a clinical perspective, our results indicate that temperament traits may provide useful information about which HR infants are less likely to develop ASD but are not useful in predicting which HR infants will develop ASD or an atypical outcome. Future studies should increase our understanding of the role of temperament when it comes to individualizing interventions. Knowledge about temperament traits that influence adaptive functioning might help to improve the benefit of interventions in young children at risk of ASD.

SUPPLEMENTAL INFORMATION

CLASSIFIERS

To predict autism at pre-diagnostic ages, we performed a classifier analysis using scores from temperamental factors as features. Seven classifiers were built based on different features: 1) surgency; 2) negative affect; 3) effortful control; 4) surgency + negative affect; 5) surgency + effortful control; 6) effortful control + negative affect; 7) surgency + negative affect + effortful control. Each of these seven classifiers was tested using the intercept and slope of the linear developmental trajectories between 8 and 24 months, and using cross-sectional measures at: 1) 8 months; 2) 14 months; 3) 24 months. Thus, a total of 28 classifiers have been tested to predict HR-ASD vs. HR-Typical + HR-Atypical, and HR-ASD + HR-Atypical vs. HR-Typical.

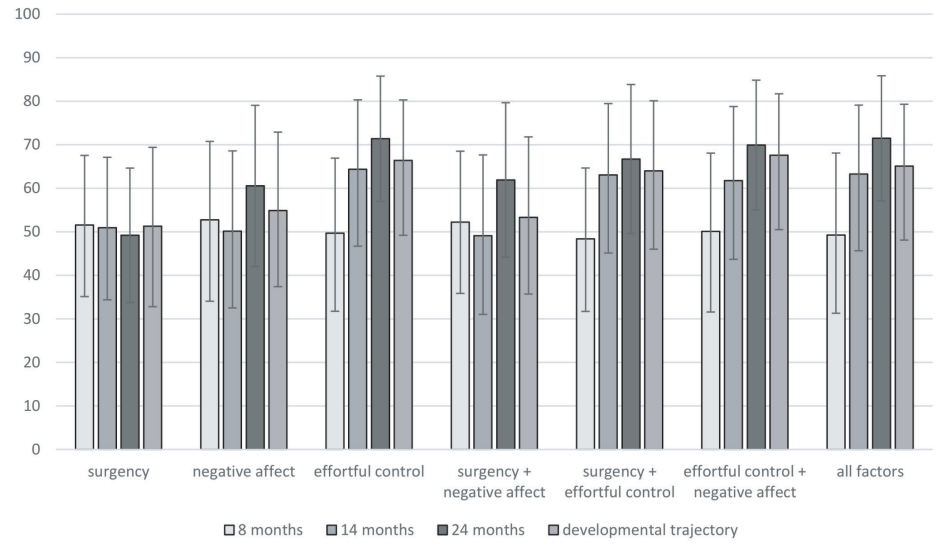


FIGURE S1. Prediction of ASD clinical outcome: AUC. In this figure the area under the curve (AUC, %) is reported for different classifiers based on temperamental factors (surgency, negative affect, effortful control) and their combinations at different time points (developmental trajectory between 8 and 24 months). Performance is computed for classification of HR-ASD from high-risk infants without a subsequent diagnosis of ASD at 36 months (i.e., HR-Typical, HR-Atypical). Individual developmental trajectories were obtained from linear modelling between 8 and 24 months, and intercept and slope have been used as features for the classifiers. 95% confidence interval is also reported for each classifier.

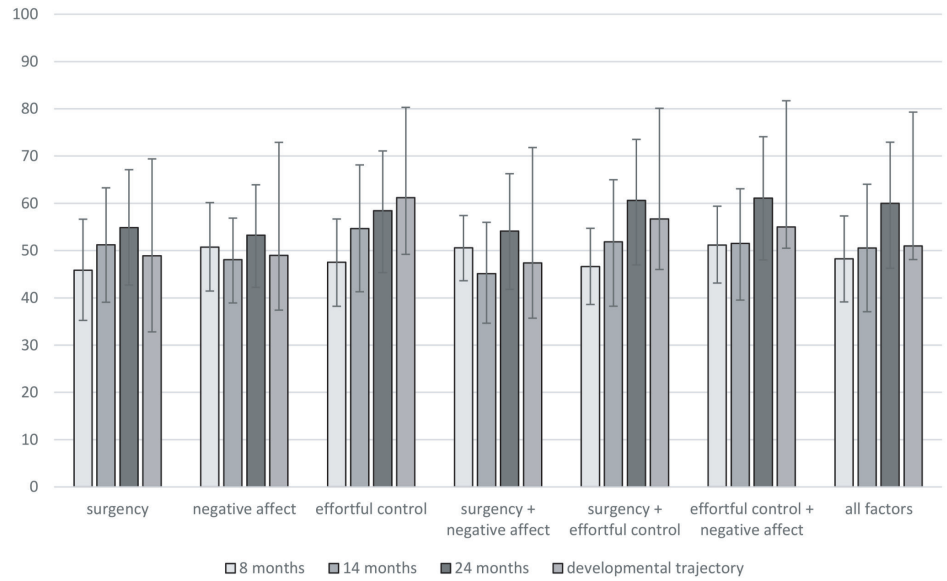


FIGURE S2. Prediction of atypical development: AUC. In this figure the area under the curve (AUC, %) is reported for different classifiers based on temperamental factors (surgency, negative affect, effortful control) and their combinations at different time points (developmental trajectory between 8 and 24 months). Performance is computed for classification of high-risk infants with atypical development at 36 months (i.e., HR-Atypical and HR-ASD) from their typically developing peers. Individual developmental trajectories were obtained from linear modelling between 8 and 24 months, and intercept and slope have been used as features for the classifiers. 95% confidence interval is also reported for each classifier.

TABLE S1. HR-ASD vs. HR-Typical + HR-Atypical. Performance metrics of temperamental factors for classifying HR siblings who later develop ASD from their peers.

	Classifier	*p*	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
14 months	Effortful control	0.047	64.4 [46.7, 80.3]	69.6 [37.4, 97.8]	59.2 [44.5, 73.4]	60.9 [47.9, 73.6]	26.6 [9.9, 45.2]	90.3 [78.3, 99.4]
24 months	Effortful control	0.006	71.4 [57.0, 82.6]	88.0 [66.2, 100]	54.8 [39.8, 69.0]	60.6 [47.0, 73.0]	29.2 [12.7, 46.5]	95.6 [87.6, 100]
	Surgency + Effortful control	0.021	66.7 [49.6, 80.8]	79.3 [49.3, 100]	54.1 [39.1, 68.8]	58.5 [44.9, 71.7]	26.7 [10.7, 44.3]	92.6 [80.8, 100]
	Effortful control + Negative affect	0.031	69.9 [55.0, 82.6]	82.6 [57.0, 98.9]	57.3 [42.4, 71.8]	61.7 [48.5, 74.7]	28.9 [12.3, 47.2]	94.0 [84.5, 99.7]
	All factors	0.020	71.5 [57.1, 83.5]	84.8 [60.5, 98.9]	58.2 [43.4, 72.6]	62.8 [49.4, 75.5]	29.9 [13.0, 48.5]	94.8 [85.7, 99.7]
Longitudinal trajectory	Effortful control	0.042	66.4 [49.2, 80.3]	79.3 [48.9, 100]	53.5 [38.9, 67.9]	57.9 [44.7, 70.6]	26.9 [11.2, 44.0]	92.3 [80.2, 100]
	Effortful control + Negative affect	0.013	67.6 [50.5, 81.7]	77.3 [47.3, 99.1]	57.8 [42.8, 72.1]	61.1 [47.9, 74.0]	28.0 [11.5, 46.2]	92.4 [81.0, 99.7]
	All factors	0.048	65.1 [48.1, 79.3]	75.2 [44.8, 97.9]	55.0 [40.2, 69.5]	58.5 [44.9, 71.7]	26.0 [9.9, 43.8]	91.5 [79.3, 99.3]

Notes. All classifiers reported in this table significantly differed from prediction at chance level (shuffle test $p<0.05$). All metrics are reported as *mean [95% confidence interval]*. 95% confidence interval was computed using bootstrap. The classifiers are divided based on the data used as features: data collected at 14 months, data collected at 24 months, intercept and slope of the longitudinal trajectory between 8 and 24 months at the individual level. The abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

TABLE S2. Best classifiers at each time point. Performance metrics for classifiers using as features temperamental factors measured at different time points for classifying HR siblings who later develop ASD (HR-ASD vs HR-Typical + HR-Atypical).

Classifier	*p*	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
Effortful control at 14 months	0.047	64.4 [46.7, 80.3]	69.6 [37.4, 97.8]	59.2 [44.5, 73.4]	60.9 [47.9, 73.6]	26.6 [9.9, 45.2]	90.3 [78.3, 99.4]
All factors at 24 months	0.020	71.5 [57.1, 83.5]	84.8 [60.5, 98.9]	58.2 [43.4, 72.6]	62.8 [49.4, 75.5]	29.9 [13.0, 48.5]	94.8 [85.7, 99.7]
Effortful control + Negative affect between 8 and 24 months	0.013	67.6 [50.5, 81.7]	77.3 [47.3, 99.1]	57.8 [42.8, 72.1]	61.1 [47.9, 74.0]	28.0 [11.5, 46.2]	92.4 [81.0, 99.7]

Notes. Classifiers for HR atypically developing siblings (HR-ASD + HR-Atypical vs HR-Typical) were not included due to performance not significantly different from chance level. All classifiers reported in this table significantly differed from prediction at chance level (shuffle test $p<0.05$). Decision on the best classifier was based on having the highest AUC within an observation time point. All metrics are reported as *mean [95% confidence interval].* 95% confidence interval was computed using bootstrap. Abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

* Comparing performance of the best classifiers at different time points via Wilcoxon tests, we found that the integrated classifier at 24 months was marginally different from effortful control at 14 months after Bonferroni correction (z=-2.14, p=0.032 with $\alpha_{Bonferroni}$=0.05/3=0.017), but not from classifier built on the developmental trajectory of effortful control plus negative affect between 8 and 24 months (z=1.86, p=0.06 with $\alpha_{Bonferroni}$=0.05/3=0.017). Classifiers at 14 months and on the developmental trajectory between 8 and 24 months were not significantly different (z=-1.17, p=0.24 with $\alpha_{Bonferroni}$=0.05/3=0.017).

TABLE S3. Difference between classifier performance: HR-ASD vs HR-Atypical + HR-Typical. Significant differences in performance between the best classifier at a specific time point and the other classifier within the same time point were tested by a two-sided Wilcoxon rank sum test when Friedman test on all classifiers performance at each time point was significant.

Paired classifiers at 14 months (effortful control vs.)	*z*	*p*
Surgency*	-2.70	0.007
Negative affect*	-2.80	0.005
Surgency + negative affect*	-2.80	0.005
Surgency + effortful control	-0.70	0.484
Effortful control + negative affect	-1.13	0.260
All factors	-0.46	0.646
Paired classifiers at 24 months (all factors vs.)	*z*	*p*
Surgency*	-2.70	0.007
Negative affect*	-2.70	0.007
Effortful control	-0.51	0.610
Surgency + negative affect	-2.50	0.013
Surgency + effortful control	-1.58	0.114
Effortful control + negative affect	-0.98	0.327
Paired classifiers on longitudinal trajectories (effortful control + negative affect vs.)	*z*	*p*
Surgency*	-2.81	0.005
Negative affect*	-2.81	0.005
Effortful control	-0.42	0.678
Surgency + negative affect*	-2.81	0.005
Surgency + effortful control	-1.28	0.202
All factors	-1.12	0.262

Notes. Bonferroni correction was used to correct post-hoc Wilcoxon tests for multiple comparison (pairs=6) and results were considered significant for $p<\alpha_{Bonferroni}=0.008$ [*]. Results from Friedman tests are $\chi^2(6)=40.6$ and $p<10^{-3}$ at 14 months; $\chi^2(6)=34.8$ and $p<10^{-3}$ at 24 months; and $\chi^2(6)=40.4$ and $p<10^{-3}$ using longitudinal trajectories between 8 and 24 months.

4

Widespread atypical neural responses to faces at 8 months predict autism spectrum disorder at 3 years

*Submitted**

Tye, C.[#],
Bussu, G.[#],
Jones, E.J.H.,
Gliga, T.,
Elsabbagh, M.,
Pasco, G.,
Johnsen, K.,
Charman, T.,
Buitelaar, J.K.[##],
Johnson, M.H.[##] & The BASIS team

*Adapted from the original form for this thesis.
[#]Tye, C. and Bussu, G. share first authorship.
[##]Buitelaar, J. K. and Johnson, M. H. share last authorship.

ABSTRACT

Atypical face processing is characteristic of emerging autism spectrum disorder (ASD) in infancy, but investigating the consistency of findings at both the group- and individual-level is critical to test utility as a predictor of later ASD. We measured event-related potentials to visual noise and faces with static gaze and dynamic gaze shifts in 8-month-old infants with (n=166) and without (n=75) familial risk for ASD, and investigated the association with ASD at 36 months. Using a two-pronged approach combining case-control analyses and machine learning techniques, findings converge to indicate that infants developing ASD show alterations in speed and amplitude of responses at early sensory and later higher-level processing stages across multiple face stimuli, consistent with wide-spread atypicalities and reflecting general neural processing differences. Diffuse alterations in face processing early in infancy represent a strong candidate biomarker for later ASD development.

INTRODUCTION

Autism spectrum disorder (ASD) is defined on the basis of social and communication impairment, restricted patterns of behaviours and interests, and sensory anomalies in early childhood[8]. While reliable diagnosis of ASD can be made by 18 months to 3 years, parental concerns are often noted earlier in retrospective reports[173]. Identification of the causal pathways leading to ASD necessitates study from early infancy, prior to the onset of behavioural symptoms. Based on a sibling recurrence rate of around 20%[27], investigation of “infant siblings” (infants with an older sibling with ASD) has accumulated over the past decade, with concomitant progress in characterising candidate neural and cognitive precursors of symptom emergence in ASD[29].

The majority of investigations of infant precursors of ASD have focused on social perception and social attention as potential origins of later emerging social, cognitive and attention features. From the first year of life, infants with later ASD demonstrate emerging atypicalities in social-communicative behaviour, such as a declining interest in human faces[37, 80, 81]. Preliminary evidence suggests that these behavioural findings are accompanied by atypical neural responses to faces as measured by event-related potentials (ERPs), which provide the resolution required to investigate different temporal stages of information processing. While some studies have suggested that low-level sensory sensitivity to faces in infancy is associated with later social development[78], atypicalities in higher-level cortical processing of faces and gaze have been reported in toddlers and children with ASD[85],[82, 84]. Infant sibling studies indicate altered responses to faces versus non-faces (objects/visual noise) and a reduced differentiation of faces that shift gaze towards versus away from the viewer from 6 months of age in infants with later ASD, as indexed by higher-level cortical responses[87, 174]. Behavioural decline in attention to faces in the first year of life appears therefore to be accompanied by atypicalities in face processing measured at the neural level at both early and later stages of processing.

In ‘social first’ theories, atypicalities in social information processing early in life result in cascading effects of reduced engagement with social stimuli, reduced opportunities for social learning, and subsequent atypical development of social cognition and communication, characteristic of ASD[175]. The extent to which reported anomalies are specific to the social domain or reflect diffuse dysfunction that broadly affects neural processing across domains is, however, unknown. In order to reliably test theoretical frameworks for emergence of ASD symptoms, it is critical that the extent of universality of social/non-social processing atypicalities across infants at familial risk for ASD is established. This can be achieved by investigating specificity of social processing anomalies to certain temporal stages of

processing and/or stimuli (e.g. faces), variation according to key stratification parameters (e.g. age, developmental ability), and importantly, by moving beyond the group-level to test generalizability of effects to the level of the individual[167]. However, while group differences have been widely reported, prediction at the level of the individual remains rare.

Here, we take a two-pronged approach to investigate 1) the reproducibility of group-level differences on the predictive value of neural processing of social stimuli for later ASD outcome with previous findings, and 2) their consistency at the individual level. We presented faces with static direct and averted gaze, faces with dynamic gaze shifts towards and away from the viewer, and phase-scrambled faces (visual noise) to 8-month-old high-risk infant siblings and low-risk controls, and investigated associations between early sensory and later cortical ERP responses (P1, N290, P400 amplitude, latency) and ASD outcome at 36 months. We combined data from our previous work (cohort 1[87]) and newly collected data (cohort 2) to maximise the power of our analysis, and aimed to identify metrics with maximum reproducibility across cohorts. We first took a group-level approach, focusing on analysis of dynamic gaze and faces versus visual noise, because they have been implicated in our previous work. Second, we used an individual-level approach independently from the group-based analysis and applied supervised machine learning to classify infants with later ASD outcome at 36 months based on ERP responses at 8-months-old, in order to test the consistency of findings across individuals. We used machine learning based on a genetic algorithm[176], which is an optimization algorithm inspired by the evolutionary process of natural selection of the fittest, to select the optimal features for classification, and to investigate whether a specific subset of brain responses to faces and visual noise best predicted ASD outcome. Based on our previous findings, we hypothesised that altered sensitivity at early and later temporal processing stages to faces versus visual noise, and dynamic gaze shifts away versus towards the infant would be associated with later ASD, and evident both at the group level and for prediction of individual outcomes.

METHODS AND MATERIALS

PARTICIPANTS

247 participants from the British Autism Study of Infant Siblings (www.basisnetwork.org) took part in the current study, across two cohorts. In cohort 1, a cohort of 54 high-risk (HR; 21 male) and 50 low-risk (LR; 21 male) participated[87]. In cohort 2, an independent cohort of 116 HR (64 male) and 27 LR (14 male) participated. LR controls were full-term infants (gestational age 38-42 weeks) recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. Along with multiple other measures collected at several time points, infants were seen for the face/gaze ERP task when they were approximately 8 months old (Table 2). Subsequently 243 were seen for assessment around their third birthday by an independent team. Four HR infants (across cohorts) did not take part in the 36-month visit and were excluded from the analysis. Two additional LR children were absent in the 36-month visit but were included in the analysis as they showed typical development at the previous visits. Of the remaining infants, exclusion criteria were applied: low number of trials (n<10), technical issues with the EEG recording and/or 100% missing ERP data. Final numbers included in the group-level analysis are shown in Table 3 (total n= 216; LR=68, HR-no ASD=115, HR-ASD=33; Table 2). Table S1 (*Supplementary Material*) describes the demographic and clinical characteristics of infants across cohort 1 and cohort 2.

TABLE 2: Demographic and clinical characteristics of participants across HR and LR subgroups

		High-risk		Low-risk	Group differences (LR, HR-no ASD, HR-ASD)
	All	***ASD***	***No ASD***		
ERP at 8 months					
Age in days (SD)	263.12 (34.31) *n*=147	257.91 (35.54) *n*=32	264.57 (33.98) *n*=115	257.91 (35.54) *n*=68	nsd
MSEL ELC	102.17 (15.85) *n*=147	97.37 (17.65) *n*=32	103.50 (15.13) *n*=115	106.49 (12.35) *n*=68	p=.017; HR-ASD<LR
N (% boys)	148 (49%)	33 (76%)	115 (42%)	68 (44%)	x^2=12.40, p=.002
Outcome at 3 years					
Age in months (SD)	38.47 (2.26) *n*=146[c]	38.22 (1.93) *n*=32	38.54 (2.35) *n*=114	38.23 (2.29) *n*=64	nsd
MSEL ELC (SD)	103.12 (24.07) *n*=145	87.94 (26.79) *n*=31	107.25 (21.62) *n*=114	116.81 (15.41) *n*=64	p<.001; HR-ASD<HR-no ASD<LR
ADOS-2 social affect (SD)[d]	4.91 (4.44) *n*=146	8.25 (5.38) *n*=32	3.97 (3.66) *n*=114	4.03 (2.92) *n*=64	p<.001; HR-ASD>HR-no ASD, LR

TABLE 2 CONTINUED.

		High-risk		Low-risk	Group differences (LR, HR-no ASD, HR-ASD)
	All	***ASD***	***No ASD***		
ADOS-2 restricted/ repetitive behaviours (SD)[d]	1.58 (1.66)	2.78 (1.86)	1.24 (1.44)	0.98 (1.18)	p<.001; HR-ASD>HR-no ASD, LR
ADOS-2 total (SD)[d]	6.49 (5.50)	11.03 (6.73)	5.21 (4.34)	5.02 (3.23)	p<.001; HR-ASD>HR-no ASD, LR
ADOS-2 CSS[d]	3.18 (2.60)	5.09 (3.12)	2.65 (2.16)	2.56 (1.78)	p<.001; HR-ASD>HR-no ASD, LR
ADI-R Social	4.12 (5.17) *n*=145	11.16 (5.64) *n*=31	2.20 (2.88) *n*=114	-	p<.001[e]; HR-ASD>HR-no ASD
ADI-R Communication	4.28 (4.86)	10.03 (5.15)	2.72 (3.39)	-	p<.001[e]; HR-ASD>HR-no ASD
ADI-R Behaviours/ Repetitive Interests	1.59 (2.37)	4.61 (2.60)	0.76 (1.46)	-	p<.001[e]; HR-ASD>HR-no ASD

[a] includes participants in intervention case-series

[b] only includes participants in randomised controlled trial

[c] 1 HR-ASD participant had incomplete MSEL

[d] item scores from the ADOS-G were used to calculate ADOS-2 totals

[e] ADI-R not administered to LR group in cohort 1; indicates statistical tests between HR groups; nsd=non-significant difference.

Abbreviations: ADI-R: Autism Diagnostic Interview-revised; ADOS – Autism Diagnostic Observation Schedule; CSS = ADOS-2 Calibrated Severity Scale; MSEL ELC = Mullen Scales for Early Learning Early Learning Composite;

TABLE 3: Valid trial numbers for participants included in ERP analysis by outcome group and by phase.

Phase	Contrast		LR	HR		
				Combined	No ASD	ASD
Phase 1	*Total sample*		n=50	n=54	n=37	n=17
	Static gaze					
	Direct	Trials	35.0	35.3	34.8	36.7
		Valid	20.8	22.6	22.1	23.1
	Averted	Trials	35.0	35.5	34.9	37.1
		Valid	20.7	23.2	22.6	23.9
	n		*32*	*32*	*22*	*10*
	Gaze shift					
	Toward	Trials	128.6	127.5	129.1	123.6
		Valid	58.7	63.0	63.6	59.2
	Away	Trials	129.1	125.9	127.4	122.2
		Valid	59.8	63.6	64.4	59.2
	n		*45*	*50*	*33*	*16*
	Face/noise					
	Face	Trials	69.0	67.1	66.5	68.4
		Valid	39.4	39.4	39.4	37.9
	Noise	Trials	46.8	45.9	45.3	47.9
		Valid	26.5	26.7	27.2	25.5
	n		*35*	*41*	*27*	*13*
Phase 2	*Total sample*		n=27	n=116	n=99	n=17
	Static gaze					
	Direct	Trials	25.6	26.5	26.4	26.5
		Valid	14.1	16.5	16.4	20.0
	Averted	Trials	25.1	26.6	26.5	26.6
		Valid	14.7	16.8	16.6	18.1
	n		*9*	*68*	*47*	*10*
	Gaze shift					
	Toward	Trials	99.7	105.4	105.1	105.7
		Valid	44.5	51.1	49.8	56.9
	Away	Trials	101.3	105.5	105.1	106.5
		Valid	43.7	50.5	49.7	54.7
	n		*23*	*102*	*82*	*17*
	Face/noise					
	Face	Trials	50.7	53.1	52.9	53.1
		Valid	24.1	28.0		32.6
	Noise	Trials	37.5	39.3	39.2	38.9
		Valid	19.3	22.2	21.7	24.8
	n		*20*	*84*	*67*	*12*

CLINICAL ASSESSMENT

A battery of clinical research measures were administered to all children at 36 months: the Autism Diagnostic Observation Schedule (ADOS)-generic[45], a semi-structured observational assessment, was used to assess current symptoms of ASD (all LR were administered module 2; 141 HR were administered module 2 and 23 HR were administered module 1 ADOS was not completed with 6 HR and 4 LR children). Total scores for social affect (SA) and restricted and repetitive behaviours (RRB) and total overall scores were computed (Table 2). The *Autism Diagnostic Interview – Revised (ADI-R*), a structured parent interview, was completed with parents of HR infants in cohort 1 and all children in cohort 2. These assessments were conducted without blindness to risk-group status by (or under the close supervision of) clinical researchers (i.e., psychologists, speech therapists) with demonstrated research-level reliability. The early learning composite score of the Mullen Scales of Early Learning (MSEL)[1] was used to obtain a standardized measure of cognitive abilities at each visit.

Experienced researchers determined the best estimate clinical outcome by reviewing all available information from visits performed (*ASD symptomatology*: ADOS; ADI-R; *adaptive functioning* Vineland Adaptive Behavior Scale-II[2]; and *development*: MSEL). Of the 148 HR participants included in this paper, 33 [22.3%] participants met criteria for ASD (hereafter HR-ASD) and the remaining 115 [77.7%] participants did not meet criteria for ASD (hereafter HR-no ASD), using ICD-10 criteria (cohort 1) or DSM-5 (cohort 2). There was a significant difference in clinical outcome per gender ($\chi^2(2) = 13.5$, $p = .001$), with more males receiving an ASD diagnosis than females (odds ratio, OR = 4.84; 95% confidence interval [CI; 1.93 to 12.1]; $p<.001$).

TASK

The ERP task was as described in Elsabbagh et al.[87]. Infants sat on their parents' laps at a 60cm distance from a computer screen. Gaze during stimulus presentation was recorded by video camera. Each trial block began with a static colourful fixation stimulus followed by a colour image of one of four female faces, with gaze directed either toward or away from the infant. In subsequent trials of the same block, the face remained on the screen but displayed three to six gaze shifts, alternating from directed toward to away from the infant. Faces were aligned with the centre of the screen with the eyes appearing at the same location as the fixation stimuli, to ensure infants were fixating the eye region. In addition to face trial blocks, during approximately one third of all blocks, infants were presented with "visual noise" stimuli. The latter were constructed from the same faces presented within the task, by randomizing the phase spectra while keeping the amplitude and colour spectra constant. Fixation stimuli, preceding the onset of face and noise stimuli, subtended approximately 1.6 x 1.6 degrees and were presented for a variable duration of 800 to 1200ms. Each trial lasted for 800ms.

ELECTROPHYSIOLOGICAL RECORDING AND ANALYSIS

EEG was recorded from a 128 channel Hydrocel Sensor Net, while infants were seated on the parent's lap in front of the stimulus screen. The reference electrode was positioned at the vertex (Cz in the conventional 10/20 system). The signal was digitized at a 500Hz sampling rate and band-pass filtered between 0.1-1000Hz.

Data were stored and analysed offline in EGI Netstation version 5.2.0.2 using the same protocol as Elsabbagh et al.[87]. Trials were retained only when infants were fixating on the centre of the screen at stimulus onset, without any gaze shifts, blinking or head movements during the 800ms segment following onset of the stimulus, using the concurrent video recording. Data were then corrected to the -200ms baseline. Following automated artifact rejection, detailed manual artifact rejection was undertaken by an experienced EEG researcher (CT), through visual inspection of individual trials, with the data from any sensor excluded if they contained artifacts. Missing data from 12 or fewer channels were interpolated. Otherwise the entire trial was rejected. Data were then rereferenced to the average.

Stimulus-locked epochs (-200 to 800ms peristimulus window) were averaged for the following trial contrasts: (1) faces (valid static (irrespective of gaze direction) vs. visual noise stimuli presented at the beginning of each presentation block); (2) static gaze (faces with direct vs. averted gaze presented at the beginning of each presentation block); and (3) dynamic gaze shifts (gaze toward vs. away from the infant, after appearance of the initial face within each presentation block).

Averages were computed for each participant in each experimental condition on a minimum of 10 trials per stimulus. Due to variable rates of presentation of each stimulus type, a different number of trials were included for each contrast. Valid trials produced by each outcome group and cohort did not differ (see Table 2 for valid trial numbers for each condition and final participant numbers entered into statistical group-based analyses). In order to provide a replication analysis, the occipito-temporal montages from Elsabbagh et al. (2012) were used (see *Supplementary* Figure S1) and corroborated with visual inspection of the grand average for each condition across the three contrasts.

GROUP-BASED ANALYSIS

A reproduction analysis was performed combining the previous and new independent sample (cohort 1 + cohort 2 participants) to examine consistency of findings and enable discovery of new effects. A repeated measures ANOVA was conducted on each ERP parameter that showed significant group effects in the original analysis[87], with contrast (face versus noise; static direct versus static averted gaze; dynamic gaze toward versus

away) as the within-subjects factor and group as the between-subjects factor. Outcome group was defined as: LR, HR-no ASD, HR-ASD. A set of analyses was run with cohort (1 versus 2) as an additional between-subjects factor and without covariates and followed up with post-hoc t-tests to compare ERP amplitude and latency of the HR-ASD group against other groups. Sidak correction was used to correct for multiple testing. Covariates (age at time of EEG acquisition, MSEL visual reception and fine motor t score at 36 months) were entered into a second round of analyses. Analyses were performed on SPSS v22 (http://www.ibm.com/analytics/us/en/technology/spss).

Between the 8- and 36-month visits, 47 (32%) of the total high-risk families in cohort 2 (included in the present analyses) took part in a randomized controlled trial (RCT) of parent-mediated intervention (Green et al. 2015) and a further 6 (4%) were enrolled in a similar non-RCT intervention (Green et al. 2013). There was no effect of recruitment (being enrolled in the intervention, regardless of treatment versus control group) into the intervention on the condition x group interaction effects on ERP parameters (see Supplementary Material). The intervention factor was therefore removed from further analysis.

INDIVIDUAL-LEVEL ANALYSIS

The main aims of this analysis are to classify among HR infants those who will later develop ASD at 36 months versus non-ASD siblings using ERP measures at 8 months, and identify the most relevant ERP measures of neural sensitivity to eye gaze for early prediction of ASD. In accordance, LR siblings were excluded from the classifier analysis since our main aim was to classify ASD among HR siblings, thus the final sample size for the individual-level analysis comprised a total of 145 high-risk siblings. Among the 145 HR infants used in the classifier analysis, 83 had complete ERP data available for every contrast while the other 62 had missing data for a total of 20.7%. At the individual level, imputation through expectation maximization was used to handle missing data, which showed a pattern of data Missing At Random (MAR).

We first performed feature selection using a genetic algorithm to extract information about the most relevant features for prediction of ASD; second, we performed SVM classification of HR-ASD vs. HR-no ASD.

Genetic Algorithm for Feature Selection

A total of 55 variables including gender, averaged ERP measures in response to each condition, and differential ERP responses for each contrast were used as features for the classifier analysis (see Table S3 for the list of features). This analysis investigated whether prediction of ASD outcome at 36 months was possible from measures of brain activity in response to social stimuli at 8 months, and differentiated between HR-ASD and HR-no ASD outcomes. We chose to include gender among features because of the significant difference in outcome per gender (see above). Each feature was standardized, and we used a genetic algorithm [183] to select the features for the classifier.

Feature selection is the process of finding the most relevant variables for the predictive model to reduce redundancy in the set of variables. Redundancy might in fact degrade accuracy, generalization and learning speed of the model[184]. The genetic algorithm is one of the most advanced algorithms for feature selection. It is a stochastic method for function optimization inspired by the evolutionary process of natural selection on genotype which inspired the algorithm, but it does not necessarily involve genetic data and can be applied to any kind of features. Starting from a collection (population) of candidate solutions (chromosomes; here sets of features) built from the available measures (gene pool; here features), the evolutionary process begins generating successive populations (generations) through mating, crossover and mutation[185]. The fitness is computed for each chromosome in each generation, and selection is based on the Darwinian principle of survival of the fittest, which in the end provides the best solution for the search problem. For reproduction, chromosomes are selected by evaluating the fitness value. Chromosomes with higher fitness have higher chance to be elected into the recombination pool.

In the present study, fitness is a measure of predictive performance of a 2-fold cross-validated SVM classifier built on the set of features under evaluation [chromosome]. We chose the Area Under the Curve (AUC) as the target value for fitness. The AUC is a measure of predictive accuracy for the model, computed as the area under the Receiver Operating Characteristic (ROC) curve, where the ROC curve is a plot of true positive rate vs. false positive rate for the model under evaluation.

Population size [n=100] and number of generations [n=200] were selected by an experienced researcher (KJ). The length of chromosomes, or number of features for the classifier, was selected based on the AUC level reached during the evolutionary process, and stability of the process assessed through visual inspection. Once selected the number of features (n=17), the evolutionary process was repeated n=100 times to investigate the variability in the feature space.

The feature set providing the highest AUC in the evolutionary process was selected as input for the subsequent classifier analysis (optimal set; see Table 1). In addition to it, the candidate solutions with highest AUC (higher than 85%) were selected and used for frequency analysis on the selected features. In fact, the selected sets of features have

nearly equal quality for classification, but the incidence frequency of each feature in the genetic evolutionary process provides an estimate of the relevance of each feature for the specific classification problem. The features with highest incidence (higher than 80%) were selected as input for subsequent classifier analysis (highest incidence set; see below), as well as task-based subsets of this highest incidence set (see Table 1).

Classifier Analysis

A support vector machine (SVM) algorithm with linear kernel was used for classification. After feature selection performed by the genetic algorithm (see above for details), we built 6 classifiers on different input sets of features: (1) the optimal set from feature selection; (2) the set of features with highest incidence (f>0.8) in the feature sets with highest performance (AUC>0.85) during repeated evolution of the genetic algorithm; (3) the features on gaze shift among the most frequent features; (4) the features on static gaze and face versus visual noise processing among the most frequent features; (5) the features on static gaze processing among the most frequent features; (6) the features on static gaze, face and visual noise processing among the most frequent features. The classifier was fully cross-validated via 10-fold cross-validation, and the sample partitioning into folds was stratified for binary outcome (i.e., ASD vs. non-ASD). All classification analyses were completed using custom scripts implemented on Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016), and the *LIBSVM* toolbox [186] was used for the SVM algorithm. To evaluate classification performance, we computed AUC, sensitivity, specificity, accuracy, negative predictive power (NPV), and positive predictive power (PPV) from the ROC curve. 95% confidence intervals (CI) for each performance metric were computed using bootstrap with n=10000 repetitions. The final metrics with errors were obtained from the average and standard deviation values over 1000 repetitions of the entire procedure, and the 95% confidence interval of each metric was also averaged over repetitions. We tested for significant difference of the classifier performance (AUC) from chance level through a shuffle test [187]. Labels in the training set were randomly shuffled, and classifiers trained to predict the shuffled random labels. Then, AUC was computed for these classifiers predicting true test labels. This procedure was repeated n=10000 times to estimate the null distribution of AUC and test whether classifiers perform significantly better than random. The *p-value* of AUC for each classifier is reported. Finally, performance of the different classifiers was compared through a nonparametric Friedman test, testing the main effect of classifier on prediction. When the Friedman test was significant, we performed post-hoc paired Wilcoxon and used Bonferroni correction to account for biasing effects due to multiple comparisons.

RESULTS

Figure 1 illustrates the flow chart of the statistical analysis.

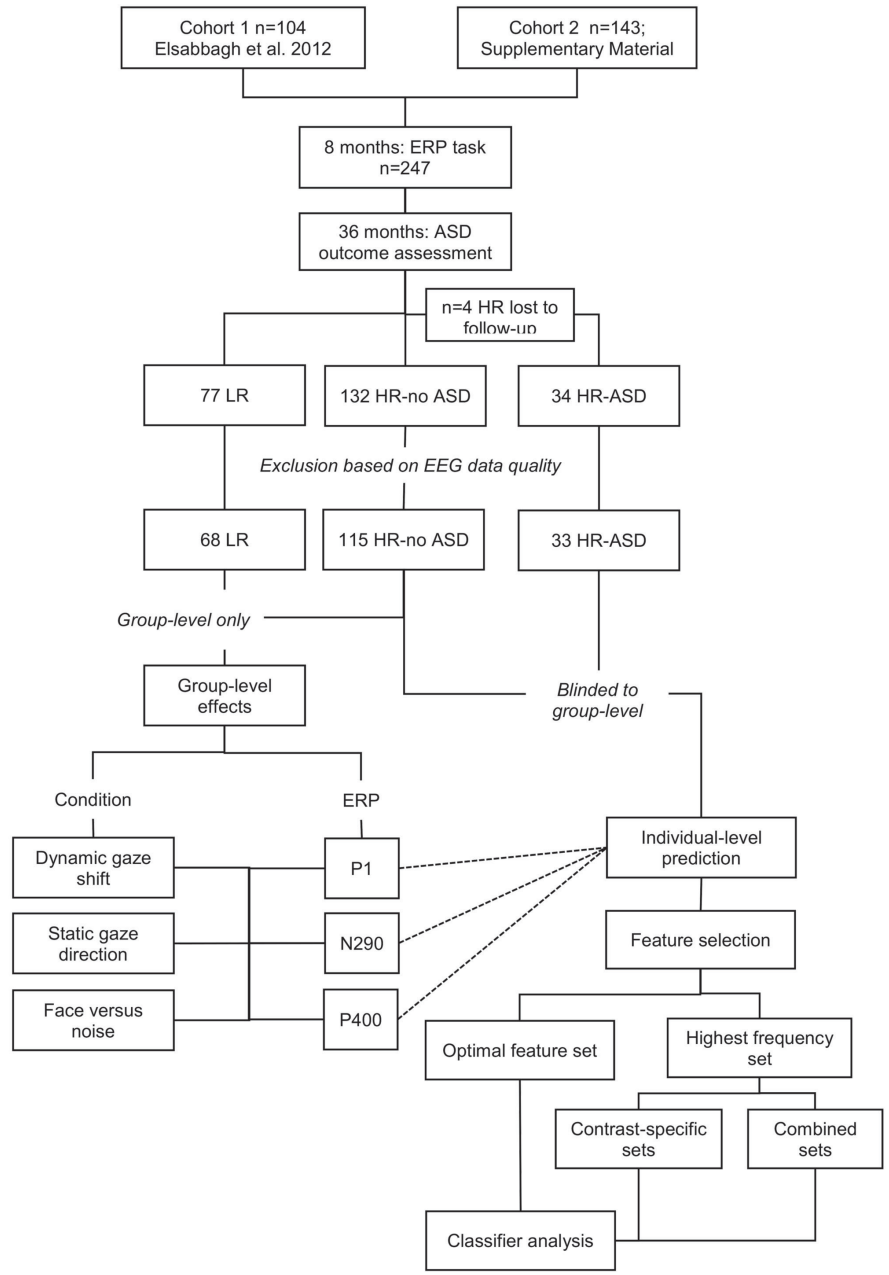


FIGURE 1: Flow chart of statistical analysis strategy

Group-level analyses were performed using repeated measures analysis of variance (ANOVA) with group [low-risk (LR); high-risk no ASD (HR-no ASD); high-risk ASD (HR-ASD)] as the independent variable with condition as the repeated measure for each contrast and ERP parameter of interest, controlling for cohort and key stratification parameters (covariates; baseline age in months, outcome non-verbal ability). Gender was not a significant covariate in any analysis and was not retained. Figure 1 illustrates the flow chart of the statistical analysis.

Next we assessed the convergence of these group-level findings with individual-level prediction of HR-ASD versus HR-no ASD classification. Low-risk controls were excluded from the classifier analysis since our aim was to predict ASD outcome among HR siblings. We used an approach that was blinded to the group-level findings but explored the same ERP components, yielding 55 features (combinations of components and conditions or differences between conditions plus gender, Table S3). Feature selection improved prediction of ASD outcome with respect to all available ERP measures (see Table S5). The genetic algorithm indicated the optimal set of features for prediction of ASD based on the optimization of AUC (see Table 1), leading to an AUC of 80.4% (95% CI, [72.4; 87.6]; p<.001). However, different evolutionary runs may result in different optimal sets of features for the classifier that complement each other and have nearly the same value for classification. To identify which ERP components in response to faces or visual noise at 8 months contribute most to predicting ASD clinical diagnoses at 36 months, we performed a frequency analysis on the features selected by the genetic algorithm providing the highest AUC during repeated evolution. Frequencies are shown in Figure 2. Of note, gender was highly relevant for classification and was selected with 100% frequency among sets providing the highest AUC. We tested the classifier for the effect of confounding variables and found that the predictive model did not depend on age and explained variance over and above gender and NVT-scores (See *Supplementary Material* for details).

Among the most relevant features, we separated features from different stimuli to test predictive power for ASD and evaluate the relevance of different tasks for prediction of ASD. Predictive performance of the different classifiers is shown in Figure 3 and detailed results are shown in Table 1.

A

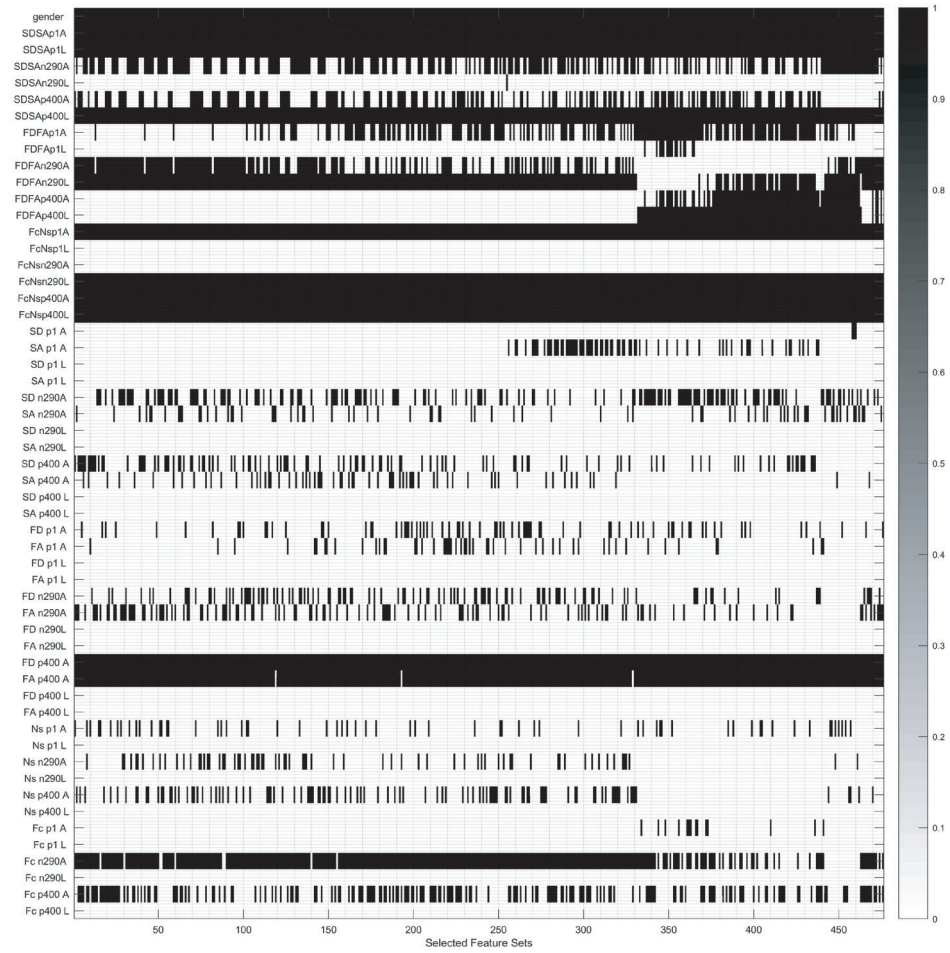


FIGURE 2: Feature selection by the genetic algorithm. ***[Legend on the next page].***

B

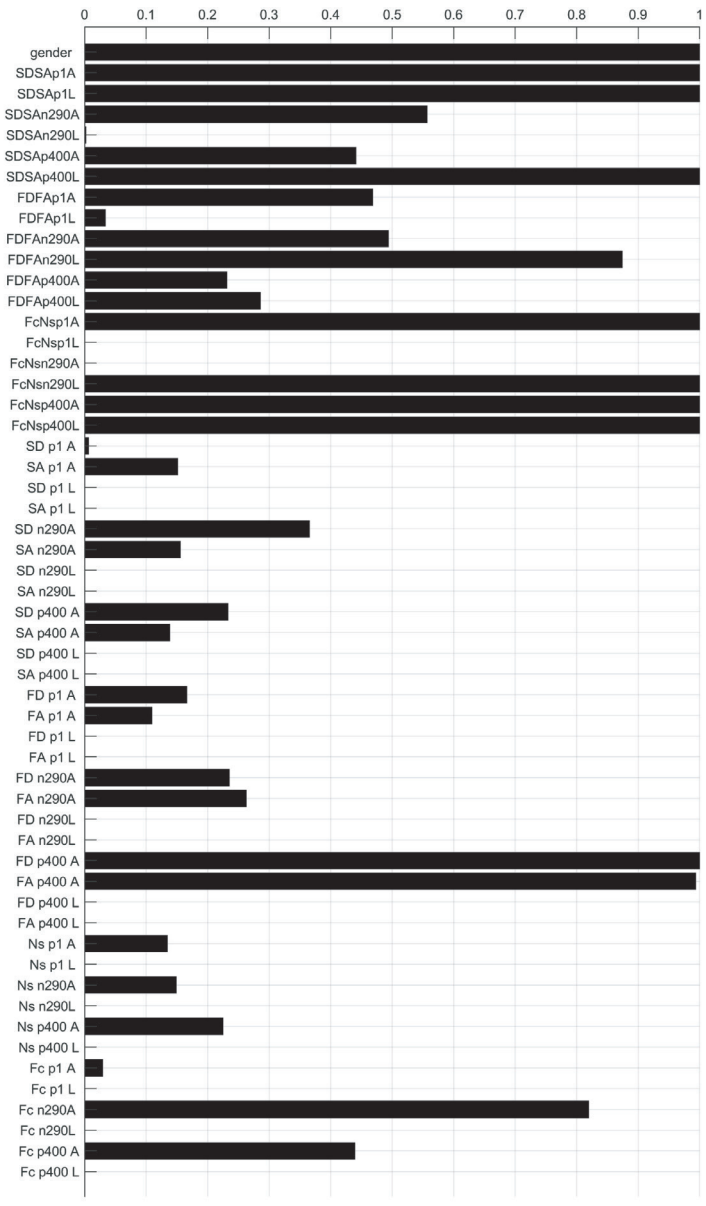


FIGURE 2: Feature selection by the genetic algorithm. This figure shows the selection occurrence (binary) of features by the genetic algorithm for the classifiers with highest AUC (higher than 85%) during repeated evolution [panel A]. Total recurrence of each feature is shown in the barplot [panel B], indicating the relevance of each feature for prediction of ASD outcome.

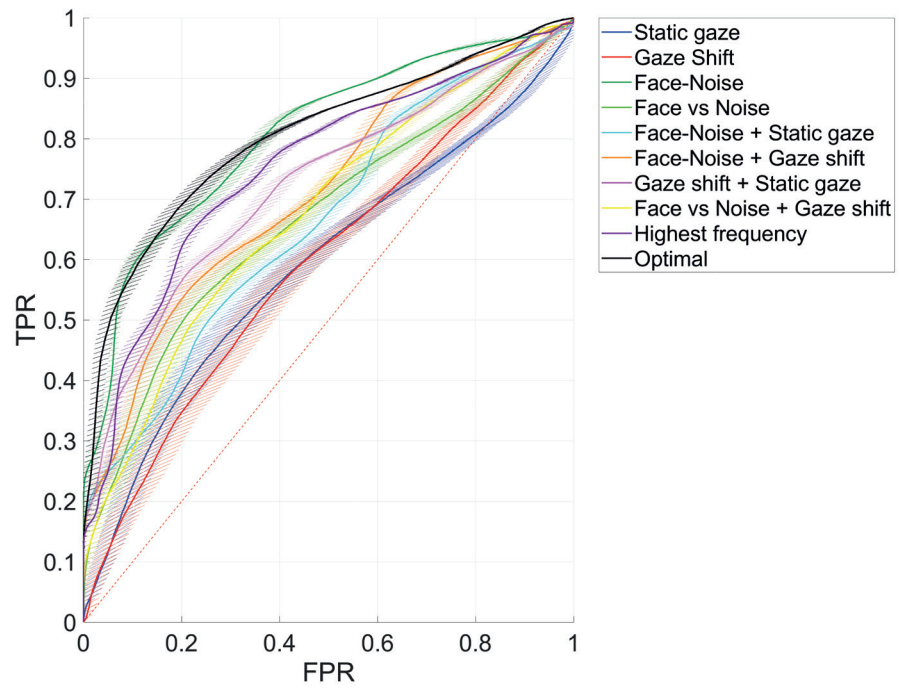


FIGURE 3. Predictive performance of different classifiers for classification of HR-ASD. This figure shows the Receiver Operating Characteristic (ROC) curve for classifiers using different set of features to classify HR-ASD among HR siblings. Each curve shows the average among 1000 repetitions of cross-validated classifications, while the shaded areas indicate one standard deviation around the mean. Random predictors result in bisecting lines as ROC curves (red dashed line), while deviations in the upper hemifield indicate an increase in predictive accuracy.

TABLE 1. Performance of different classifiers at predicting individual ASD diagnoses at 36 months.

	Optimal Set		Highest Incidence		Gaze Shift		Static Face		Face vs Noise		Face - Noise		Face vs Noise + Gaze shift		Face – Noise + Static Face		Face – Noise + Gaze shift		Gaze shift + Static Face	
	Mean	***P***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***	***Mean***	***p***
AUC	80.4 [72.4; 87.6]	<.001	75.4 [66.2; 83.8]	<.001	59.4 [51.8; 69.9]	.40	59.0 [51.8; 69.1]	.11	66.8 [57.1; 76.2]	0.004	80.7 [72.5; 88.0]	<.001	67.7 [57.5; 77.4]	.005	66.4 [56.2; 76.1]	.004	71.3 [61.5; 80.3]	<.001	72.3 [63.0; 80.9]	<.001
Accuracy	78.9 [71.8; 85.0]	<.001	72.7 [66.9; 81.9]	<.001	59.0 [57.2; 70.7]	.15	59.9 [58.3; 71.8]	.46	70.5 [61.5; 76.2]	<.001	74.9 [71.2; 84.4]	<.001	67.6 [60.4; 79.6]	.006	64.3 [60.2; 74.0]	.03	68.5 [63.2; 77.6]	0.003	70.0 [64.8; 79.2]	<.001
Sensitivity	67.3 [52.2; 85.3]	.39	73.5 [47.6; 85.2]	.28	46.0 [24.7; 88.8]	.67	35.4 [29.2; 87.4]	.78	69.0 [35.9; 78.5]	.37	65.5 [53.4; 88.7]	.42	47.8 [33.6; 85.7]	.67	56.6 [24.0; 89.9]	.53	55.8 [38.9; 29.3]	0.56	74.3 [39.9; 81.6]	0.29
Specificity	90.6 [69.3; 99.5]	.05	71.9 [59.8; 96.8]	.29	71.9 [33.3; 97.3]	.32	84.4 [37.2; 96.8]	.20	71.9 [55.8; 97.4]	.32	84.3 [64.3; 99.5]	.15	87.5 [45.7; 96.9]	.10	71.9 [39.7; 99.8]	.29	81.3 [47.6; 97.9]	0.17	65.6 [59.5; 98.5]	0.41
PPV	87.8 [72.2; 99.3]	.01	72.3 [66.2; 94.8]	.08	62.1 [56.2; 92.9]	.34	69.4 [57.4; 92.9]	.24	71.1 [62.4; 95.3]	.12	80.7 [70.3; 99.4]	.04	79.3 [59.7; 93.8]	.03	66.8 [58.9; 99.5]	.15	74.8 [62.1; 96.4]	0.05	68.4 [65.2; 97.4]	0.17
NPV	73.5 [66.4; 84.2]	.07	73.0 [63.4; 82.4]	.07	57.1 [55.1; 79.1]	.62	56.6 [56.1; 81.1]	.75	79.3 [59.74; 93.8]	.15	71.0 [66.4; 86.5]	.11	62.6 [57.6; 79.0]	.30	62.4 [56.4; 82.3]	.29	67.7 [59.8; 83.7]	0.22	71.9 [60.7; 78.9]	0.11

[Abbreviations on the next page]

This table shows performance metrics of classifiers for different input sets of features discriminating HR sibling who developed ASD from those who did not (*HR-ASD vs HR-Atypical + HR-Typical*). The significance of classification AUC was determined by permutation test, the resulting *p-values* are reported. Prediction was considered significant if $p<0.05$. 95% bootstrap confidence interval (CI) is reported in parentheses. All metrics are reported as *mean [CI]*.

Abbreviations: AUC = area under the curve; PPV = positive predictive power; NPV = negative predictive power.

Optimal set: gender; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency; SD-SA n290 amplitude; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; FD-FA n290 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude; FD-FA n290 amplitude; SD p4 amplitude; FD n290 amplitude; F p4 amplitude.

Highest incidence set: gender; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; FD-FA n290 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude.

Gaze shift set: gender; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency.

Static face set: gender; FD-FA n290 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude.

Face vs Noise set: gender; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency.

Face - Noise set: gender; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude.

Face vs Noise + gaze shift set: gender; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency.

Face - Noise + static face set: gender; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; FD-FA n290 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude.

Face - Noise + gaze shift set: gender; N-F p1 amplitude; N-F n290 latency; N-F p4 amplitude; N-F p4 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency.

Gaze shift + static face set: gender; SD-SA p1 amplitude; SD-SA p1 latency; SD-SA p4 latency; FD-FA n290 latency; FD p4 amplitude; FA p4 amplitude; F n290 amplitude.

Here, we report findings for each contrast across group and individual levels, and the overall individual prediction findings. See Figure 4 for grand average ERPs to face stimuli for each group and contrast, means and standard errors of amplitude and latency of the three face-sensitive ERPs in each subgroup for each contrast.

ALTERED SPEED OF RESPONSES TO FACES VERSUS VISUAL NOISE PREDICTS LATER ASD

At the group level, there was a significant condition x outcome interaction on N290 latency in the face/noise contrast ($F(2,171)=3.61$, $p=.029$), whereby the HR-ASD group did not show a stimulus differentiation, while the LR ($p=.010$, $d=0.61$) and HR-no ASD ($p=.021$, $d=0.52$) groups showed longer latency to faces compared to noise, with no difference between LR and HR-no ASD groups ($p=.551$, $d=0.10$). This did not vary by cohort ($F(1,170)=0.76$, $p=.386$) and neither of the covariates had a significant interaction (ps>.41). There were no significant condition x outcome group interactions on P1 amplitude ($p=.24$), P1 latency ($p=.319$), N290 amplitude ($p=.79$), P400 amplitude ($p=.68$) or P400 latency ($p=.11$). At the individual level, N290 latency to visual noise versus faces was also selected as a relevant feature for ASD

outcome (incidence higher than 80%; Figure 2), in addition to P1 amplitude, and P400 amplitude and latency in response to faces versus visual noise. Classifier analysis revealed the face versus noise contrast showed moderate predictive power for ASD outcome (AUC=66.8% [95% CI: 57.1; 76.2], p=.004; see Table 1 for combination of features).

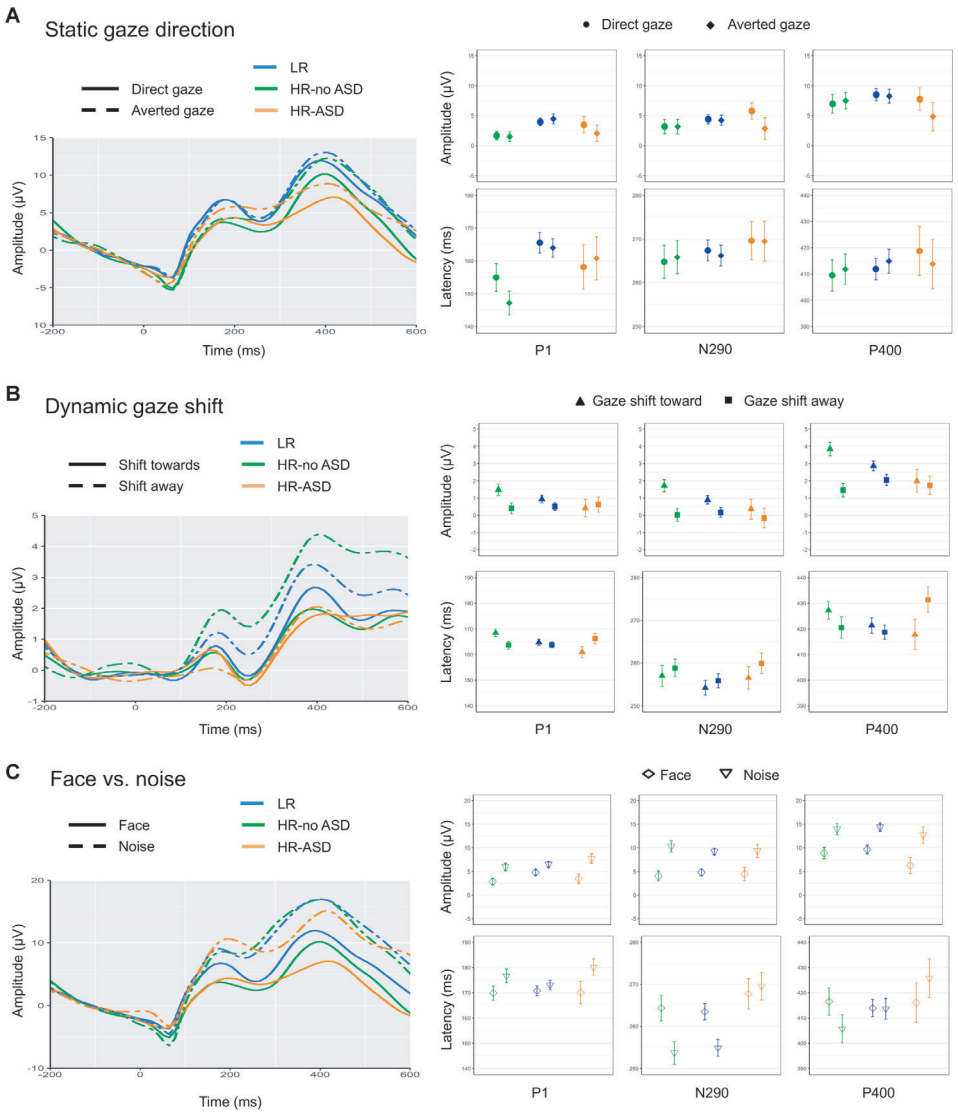


FIGURE 4: Group findings. This figure illustrates grand average ERPs [left column] and means and standard errors of amplitude and latency for ERP parameters in each group [right column] for each contrast: (A) direct vs averted static gaze; (B) gaze shifts directed away vs toward the infant; (C) face vs visual noise.

ALTERED PROCESSING OF DYNAMIC GAZE SHIFTS

Our previous work indicated infants with later ASD outcome as a group showed altered processing of dynamic gaze shifts[87]. Accordingly, at the group level, there was a significant condition x outcome group interaction on P1 latency (F (2, 213)=4.95, p=.008), whereby the HR-ASD group had longer latency to gaze shifting away versus towards compared to longer latency to gaze shifting towards versus away in the LR (p=.002, d=0.74) and HR-no ASD groups (p=.047, d=0.43), with no difference between LR and HR-no ASD (p=.122, d=0.26). There was no significant interaction with cohort (F (2,212)=1.48, p=.226), but the condition by outcome interaction became only a trend when baseline age and outcome non-verbal ability were entered as covariates (F(2,201) = 2.45, p=.089). Further analysis indicated that lower non-verbal ability was associated with longer P1 latency to gaze shifting towards versus away (r=-.18, p=.008), with no association with age (r=.07, p=.32). Although there was no condition x outcome interaction effect on P1 amplitude (p=.147) at the group level, feature selection showed P1 amplitude and latency to gaze shifts directed towards the infant versus away among the most relevant features for prediction of ASD at the individual level (Figure 2).

There was also a significant condition x group interaction on P400 amplitude in the dynamic gaze contrast (F (2,212)=4.13, p=.017), reflecting enhanced amplitude to gaze shifting away versus towards in the LR group, compared to enhanced amplitude to gaze shifting towards versus away in the HR-ASD group (p=.016, d=0.46) and HR-no ASD group (p=.014, d=0.41), with no difference between HR groups (p=.482, d=0.12). There was no interaction with cohort (F (1,211) = 0.27, p=.605) and the interaction was retained when covariates were entered (F (2,199)=3.08, p=.048), with no interaction with covariates (ps>.19). Consistently with the lack of differentiation between HR groups, P400 amplitude was not selected as a relevant feature for prediction of ASD among HR siblings at the individual level.

There were also group differences in P400 latency in the dynamic gaze shift comparison (F(2, 208)= 3.51, p=.032). The HR-ASD group showed longer P400 latency to gaze shifting towards versus away from the viewer, with an opposite effect in LR (p=.011, d=0.55) and HR-no ASD (p=.021, d=0.47), and no significant difference between LR and HR-no ASD (p=.572, d=0.10). This did not vary by cohort (F(2,207)=0.91, p=.342) and was not influenced by covariates (ps>.24). Post-hoc t-tests revealed significant differences between HR-ASD and LR (p=.009, d=0.60) and HR-no ASD (p=.019, d=0.47), but not between LR and HR-no ASD (p=.488, d=0.12). Individual-level feature selection also indicated the relevance of P400 latency. There were no group effects on N290 amplitude (p=.190) or N290 latency (p=.858), and these were not selected as relevant features at the individual level for the dynamic gaze condition. Overall, classifier analysis showed poor predictive power of

neural responses to dynamic gaze shifts (AUC=59.4% [95% CI:51.8; 69.9], p=.15).

Since we observed alterations in dynamic gaze processing across both P1 and P400 components, we sought to test whether these represented manifestations of the same or different underlying phenomena. No significant associations between latency of the P1 and P400 ERP difference scores for gaze shift towards versus away were found (whole sample: r=0.3, p=.697; LR: r=-.05, p=.648; HR-no ASD: r=.05, p=.622; HR-ASD: r=-.08, p=.664). A hierarchical regression indicated independent effects of the P1 and P400 difference scores on outcome group; the P1 latency difference remained after entering the P400 difference (beta=.211, p=.002) and the P400 latency difference remained after entering the P1 difference (beta=.160, p=.017).

NEURAL PROCESSING OF STATIC GAZE DIRECTION

At the group level, there were no significant condition x outcome group interactions on ERP parameters for the static gaze (direct versus averted) contrast, for P1 amplitude (p=.446) and latency (p=.252), N290 amplitude (p=225) and latency (p=.828), and P400 amplitude (p=.178) and latency (p=.515), supporting the specificity of our findings to dynamic gaze shifts. However, responses to static gaze direction were selected among the relevant features for individual prediction of ASD outcome. Specifically, feature selection pointed to P400 amplitude in response to direct and averted gaze, with the HR-ASD group showing reduced amplitude to both, and N290 latency to direct versus averted gaze, with longer latency to direct compared to averted gaze in the HR-ASD and the opposite effect in the HR-no ASD group. Furthermore, N290 amplitude to faces with static gaze (an average between direct and averted gaze) was selected as a relevant feature for individual-level prediction, with lower amplitude in the HR-ASD group compared to HR-no ASD. However, classifier analysis showed poor predictive power of neural responses to static gaze (AUC=59.0% [95% CI: 51.8; 69.1], p=.11), which predicted at chance level.

COMBINATION OF FEATURES IMPROVES INDIVIDUAL-LEVEL PREDICTION OF ASD OUTCOME

The highest incidence set (Table 1) indicated the most relevant features for prediction of ASD at the individual level and included a pattern of early and late components in response to the different face stimuli (reported separately for each contrast above), from which classification was possible with 75.4% AUC (95% CI: [66.2, 83.8], p<.001). The optimal set selected by the genetic algorithm outperformed prediction from the most relevant features (p<.001), showing that there is additional useful information in the complete optimal set for prediction of ASD not explained by the features with highest incidence.

Although the evaluation of classifier performance for contrast-specific subsets of features showed largely poor predictive power, combined sets of brain responses to different contrasts improved predictive accuracy compared to single contrasts, as shown by the integration of responses to the face vs. noise contrast and separate responses to faces with directed or averted static gaze (Face-Noise feature set: AUC=80.7% [95% CI: 72.5; 88.0], p<.001), the integration of responses to the face vs. noise and gaze shift contrasts plus separate responses to faces with directed or averted static gaze (Face-Noise + Gaze shift set: AUC=71.3% [95% CI: 61.5; 80.3], p<.001), and the integration of responses to the static and dynamic gaze contrasts (Gaze shift + Static Face: AUC=72.3% [95% CI: 63.0; 80.9], p<.001). On the other hand, the addition of the static gaze contrast to the Face-Noise set drops predictive power to 66.4% AUC (Face-Noise + Static Face: 95% CI: [56.2; 76.1], p=.004), while the integration of responses to the face vs noise and the gaze shift contrasts only yielded to a 67.7% AUC (Face vs Noise + Gaze shift: 95% CI: [57.5; 77.4], p=.005). The non-parametric Friedman test showed a significant difference in performance of the different classifiers ($\chi^2(9) = 8229$, p<.001), and the post-hoc paired Wilcoxon tests showed significant differences among all pairs (Bonferroni corrected p<.001), except for the static gaze set vs. the gaze shift set (z=-2.85, Bonferroni corrected p=.19). Overall, looking at AUC, the best performing classifier at predicting 36-month ASD diagnosis was the classifier built on the Face-Noise stimulus set.

DISCUSSION

We investigated the reproducibility of group-level predictions for ASD outcome based on face processing, and the consistency of these as individual outcome predictors. Our results indicate alterations in both early sensory and later higher-level stages of neural processing of social (face/gaze) and non-social (noise) stimuli that differentiate infants with later ASD from those without later ASD, and that are robust across both the group- and individual-level. These findings support a theoretical framework in which diffuse anomalies in ASD reflect broader general differences in neural processing from an early age.

COMBINATION OF ATYPICALITIES IN NEURAL PROCESSING OF FACES ACROSS THE TIME-COURSE OF FACE PROCESSING PREDICTS ASD OUTCOME

Group-level and individual-level analyses converged to suggest that the socially relevant processes of detecting a face and detecting a shift in gaze are altered across a long time-course of information processing from the shortest latency components, and across multiple stimuli manipulations. Infants with later ASD (HR-ASD) tend to show as a group longer P1 and P400 latency to gaze shifting towards versus away from the viewer, with the opposite finding in infants without later ASD (LR and HR-no ASD). While the same effect is shown at early and later stages of processing, there was no direct association between P1 and P400 latency to dynamic gaze shifts. In combination with the effect of age and non-verbal ability specifically on P1 latency only, this suggests that the P400 difference is not directly attributable to inputs from earlier stage processing. In addition, longer N290 latency to faces compared to visual noise shown in the LR and HR-no ASD groups was diminished for the HR-ASD group. The more rapid peaking of the N290 to faces over noise in ASD is similar to previous findings on the P400, and may suggest reduced attention engagement to faces (and social stimuli more generally) as the neural resources allocated to process a face peak and decline more rapidly[174]. The same measures were found to be part of a broader pattern predicting later ASD diagnosis among high-risk siblings at an individual level with approximately 80% accuracy. This pattern includes shorter and longer latency components in response to different face manipulations, consistent with widespread atypicalities in neural processing across processing stages and across the high-risk population. Initial differences in these neural processing stages may subsequently trigger a cascade of events that result in symptoms characteristic of ASD[70].

We also partially reproduced our previous findings from cohort 1 with the larger combined sample; P400 amplitude to gaze shifting away from the infant, versus towards, was shown in the LR infants with a diminished effect of gaze shifting in both HR groups. This may explain why P400 amplitude in the dynamic gaze contrast was not an optimal feature for outcome classification in individual-level analyses. Reduced P400 differentiation of dynamic gaze in cohort 1 has recently been shown to predict ASD outcome in middle childhood, both in terms of stable diagnosis from 3 to 7 years and a 'late' diagnosis made after 3 years of age[177]. Follow-up of the larger cohort will help to determine whether this particular feature is only detected in a subgroup of individuals or is more strongly related to longer-term outcome.

GENERALIZABILITY ACROSS INDIVIDUALS

Findings at the group-level appear to generalize across cohorts since there were no significant interaction effects with cohort. This suggests relative robustness of the observed effect of outcome on latency of cortical processing, in particular for dynamic gaze shifts over static gaze processing[87]. However, group differences in changes to P1 latency in response to gaze shift direction were not retained when age at ERP assessment and non-verbal ability at outcome assessment were entered. Further analyses indicated lower non-verbal ability was associated with longer P1 latency to gaze shifting towards versus away, which highlights a potential source of heterogeneity across individuals.

Group effects, which report on the average HR-ASD sibling, do however disguise inter-individual variations and may not inform on the different underlying cognitive mechanisms preceding ASD. To test pervasiveness of difficulties in face processing across high-risk siblings developing ASD, we assessed predictive value of early neural sensitivity to different face stimuli for ASD outcome at an individual level through the AUC of the corresponding classifier. The AUC is an effective and combined measure of sensitivity and specificity to test the inherent ability of the predictor, providing a useful metric to evaluate diffusivity of the predictive features within the examined population[178]. In fact, a good class separation between infants with and without later ASD diagnosis can be obtained only if the predictive features are both sensitive, thus diffuse across the ASD group, and specific to the ASD group.

A broad pattern of alterations across the time-course of neural processing contributed to prediction at the individual level, rather than single ERP components. Furthermore, task-specific SVM classifiers built on subsets of the most relevant features for prediction of ASD allowed the assessment of predictive power from different stimuli. Results showed that differential responses to the face vs. noise contrast integrated with separate ERP responses to faces with static gaze had the highest predictive power for ASD (80.7% AUC; 95% CI [72.5; 88.0]) while predictive power for neural responses to gaze shifts alone was poor (59.4% AUC; 95% CI [51.8; 69.9]). This suggests a widespread main effect of stimulus difference in social content versus non-social content for discriminating infants who go

on to develop ASD from those who don't, while an overlap between groups in individual variation for neural sensitivity to dynamic gaze suggests that not all individuals in the ASD class deviate in dynamic gaze processing compared to typical responses. Further work should investigate whether atypical responses to dynamic gaze shifts map on to different profiles of atypical gaze later and define a meaningful subtype of the ASD phenotype[167].

CONCLUSIONS

Research on infants at high-risk for ASD provides supporting evidence for intervention at an early age, when the core symptoms of ASD have not emerged yet[90, 133, 179]. Still, early intervention studies are conducted at the group-level, without information on individual probabilities of developing ASD beforehand (except for reported recurrence rates of ASD among HR siblings)[27]. Thus, individual prediction of ASD in the first year of life might be crucial to identify the infants who need intervention and enable early-targeted intervention. We significantly reduced the age of detection compared to previous classification studies on behavioural measures[55, 113, 114] and ERP parameters[180]. Other studies already showed that prediction of ASD in the first year of life is possible using functional and structural magnetic resonance imaging[111, 112], with more than 94% accuracy. However, our results extend these findings by using ERPs to predict a more stable ASD diagnosis at 36 months in a larger HR sample[111]. Furthermore, ERPs represent a more cost-effective, mobile, and infant-friendly neuroimaging technology, providing potential utility for early screening and inclusion as proxy outcome markers for intervention trials[181]. To test this, future work should determine whether these parameters are sensitive to the effects of early intervention[182], together with tests of integration with other risk markers (e.g. genetic factors, multimodal MRI, parent-child interaction and other infant behavioural measures) to improve individual prediction of ASD outcome.

It is important to consider methodological limitations in machine learning related to relatively small sample size and model reliability[105]. Thus, generalizability of the identified model must be tested through external validation. Of note, we chose to focus on the most relevant features based on incidence of feature selection (although the AUC was higher for the optimal feature set ($p<.001$)), because the most relevant features are more likely to be selected during validation in an independent sample, while the optimal set might change.

Our two-pronged approach represents the first attempt to investigate robustness and generalizability of group-level factors across infants at-risk for ASD, showing a diffuse pattern of atypicalities across the whole time course of neural processing in response to different face stimuli. This adds to the literature illustrating early structural atypicalities[112] by showing early diffuse functional atypicalities. The findings indicate that robust differences in early sensory and later cortical stages of face processing in the first year of life can predict individual ASD outcome at 3 years. A focus on these very early mechanisms allows development of more targeted pre-diagnostic interventions to infants who are at the highest risk for developing later impairments.

SUPPLEMENTARY MATERIAL

CONTENTS:

Table S1: Demographic and clinical characteristics by cohort
Table S2: Group x condition interactions p-value for each contrast
Table S3: Features for the classification problem
Table S4: Prediction of gender and non-verbal t scores at 36 months
Table S5: Prediction of ASD diagnosis using all ERP variables
Table S6: Proportion of participants who entered intervention or received treatment in cohort 2
Analysis S1: Analysis accounting for participation in early intervention
Analysis S2: Cohort 2-only analysis
Analysis S3: Confounding variables in the individual-level analysis
Figure S1: Selected channel montages

TABLE S1: Demographic and clinical characteristics by cohort (participants included in analyses)

	Cohort 1 (n=94)	Cohort 2 (n=122)	Group difference
Male n (%)	42 (40.38%)	78 (54.54%)	$x^2 = 4.83$, p=.028
ERP at 8 months			
Age in days (SD)	239.39 (37.66)	276.53 (25.76)	t (155) = -8.17, p<.001
MSEL ELC (SD)	98.90 (12.95)	107.07 (15.44)	t (213) = -4.11, p<.001
Outcome at 3 years			
Age in months (SD)	37.87 (2.81)	38.81 (1.62)	t (137) = -2.84, p=.005
MSEL ELC (SD)	110.16 (19.43)	105.11 (24.71)	nsd
ADOS-2 social affect (SD)	5.86 (3.87)	3.69 (3.96)	t (208) = 3.97, p<.001
ADOS-2 RRB (SD)	1.47 (1.63)	1.34 (1.50)	nsd
ADOS-2 CSS (SD)	3.62 (2.44)	2.51 (2.25)	t (208) = 3.41, p=.001
ADI-R Social[a]	4.79 (5.45)	3.26 (4.72)	nsd
ADI-R Communication[a]	4.56 (4.93)	3.49 (4.63)	nsd
ADI-R Behaviours/Repetitive Interests[a]	1.60 (2.02)	1.31 (2.36)	nsd

Abbreviations: ADI-R: Autism Diagnostic Interview-revised; ADOS – Autism Diagnostic Observation Schedule; CSS = ADOS-2 Calibrated Severity Scale; MSEL ELC = Mullen Scales for Early Learning Early Learning Composite
[a]ADI-R not administered to LR group in cohort 1; indicates statistical tests between HR groups; nsd=non-significant difference.

TABLE S2: Group x condition interactions p-value for each contrast (no covariates)

	P1		N290		P400	
	Amp	**Lat**	**Amp**	**Lat**	**Amp**	**Lat**
Static gaze condition x outcome group	.45	.25	.23	.83	.18	.52
Low-risk: direct vs averted						
No ASD: direct vs averted						
ASD: direct vs averted						
Dynamic gaze condition x outcome group	.15	**.01***	.35	.84	**.03***	**.03***
Low-risk: toward vs away		.01*			<.001*	.22
No ASD: toward vs away		.54			.03*	.46
ASD: toward vs away		.02*			.80	.03*
Face condition x outcome group	.10	.25	.37	**.03***	.66	.14
Low-risk: face vs noise				<.001*		
No ASD: face vs noise				<.001*		
ASD: face vs noise				.58		

*p<.05

TABLE S3. Features for the classification problem.

SD p1 amplitude	SD p1 latency	FD p1 amplitude	FD p1 latency	N p1 amplitude	N p1 latency
SA p1 amplitude	SA p1 latency	FA p1 amplitude	FA p1 latency	F p1 amplitude	F p1 latency
SD n290 amplitude	SD n290 latency	FD n290 amplitude	FD n290 latency	N n290 amplitude	N n290 latency
SA n290 amplitude	SA n290 latency	FA n290 amplitude	FA n290 latency	F n290 amplitude	F n290 latency
SD p4 amplitude	SD p4 latency	FD p4 amplitude	FD p4 latency	N p4 amplitude	N p4 latency
SA p4 amplitude	SA p4 latency	FA p4 amplitude	FA p4 latency	F p4 amplitude	F p4 latency
SD-SA p1 amplitude	SD-SA p1 latency	FD-FA p1 amplitude	FD-FA p1 latency	N-F p1 amplitude	N-F p1 latency
SD-SA n290 amplitude	SD-SA n290 latency	FD-FA n290 amplitude	FD-FA n290 latency	N-F n290 amplitude	N-F n290 latency
SD-SA p4 amplitude	SD-SA p4 latency	FD-FA p4 amplitude	FD-FA p4 latency	N-F p4 amplitude	N-F p4 latency
Gender					

This table shows the list of 54 ERP features + gender (n=55 in total) used as input features for the genetic algorithm to select features for classification. *Abbreviations:* SD = gaze shift directed towards the infant; SA = gaze shift directed away from the infant; FD = static face with direct gaze towards the infant; FA = static face with gaze directed away from the infant; F = static face (average direct and averted gaze); N = visual noise.

TABLE S4. Prediction of gender and NVT-scores at 36-months

	Gender		Visual Reception Skills		Fine Motor Skills	
	Mean	***p***	***Mean***	***p***	***Mean***	***p***
AUC	52.8 [50.2; 62.5]	0.56	63.8* [51.0; 80.9]	0.07	57.6 [50.8; 68.6]	0.20
Accuracy	57.5 [54.9; 64.5]	0.30	70.3 [59.9; 81.4]	0.02	61.0 [56.9; 69.6]	0.21
Sensitivity	27.4 [20.1; 95.1]	0.90	65.6 [36.7; 94.4]	0.50	74.2 [28.8; 97.5]	0.29
Specificity	87.5 [19.4; 94.7]	0.08	75.0 [35.1; 97.6]	0.30	47.7 [24.1; 91.9]	0.65
PPV	68.7 [53.4; 82.1]	0.06	72.4 [57.7; 96.1]	0.22	58.7 [54.9; 82.0]	0.53
NPV	54.7 [53.5; 82.9]	0.75	68.6 [58.1; 89.8]	0.33	64.9 [55.5; 93.2]	0.25

This table shows performance metrics of classifiers built on the optimal set of features selected by the genetic algorithm for prediction of ASD, applied here to classify gender and differentiate HR siblings with high and low NVT scores at 36 months. Significance of classification AUC was determined by permutation test, the resulting *p-values* are reported. Prediction was considered significant if $p<0.05$. 95% bootstrap confidence interval is reported in parentheses. All metrics are reported as *mean [lower level CI, upper level CI]*. Abbreviations: AUC = area under the curve; PPV = positive predictive power; NPV = negative predictive power.

* Significant difference from classification of HR-ASD

TABLE S5. Prediction of 36-months ASD diagnoses from all available ERP variables

	Complete Set	
	Mean	***p***
AUC	60.8* [51.5; 71.6]	<.001
Accuracy	62.9 [57.2; 71.9]	<.001
Sensitivity	54.0 [28.9; 90.2]	0.51
Specificity	71.9 [32.2; 94.4]	0.27
PPV	65.8 [55.9; 88.3]	0.09
NPV	61.0 [55.5; 81.1]	0.17

This table shows performance metrics of the classifier built on the complete set of features to discriminate HR sibling who developed ASD from those who did not (*HR-ASD vs HR-Atypical + HR-Typical*). The complete set included gender, all ERP variables separately and as contrasts (i.e. shift direct vs. shift averted, face vs. noise, direct vs. averted static gaze). Significance of classification AUC was determined by permutation test, the resulting *p-values* are reported. Prediction was considered significant if $p<0.05$. 95% bootstrap confidence interval is reported in parentheses. All metrics are reported as *mean [lower level CI, upper level CI]*. Abbreviations: AUC = area under the curve; PPV = positive predictive power; NPV = negative predictive power.

* Significant difference from the optimal set and the highest incidence set (Wilcoxon z=38.7, p<.001)

SUPPLEMENTARY ANALYSIS

1) ANALYSIS ACCOUNTING FOR PARTICIPATION IN EARLY INTERVENTION

TABLE S6: Proportion of participants who entered intervention or received treatment, within cohort 2 included in this analysis (including case-series)

	HR	HR-ASD	HR-no ASD
Recruited for intervention	50 (51%)	7 (41%)	43 (52%)
Treated in intervention	27 (27%)	4 (24%)	23 (28%)

Analysis focused on significant group x condition interactions. For all ANOVAs, we included two binary terms "*treatment*" (treated versus not treated) and "*recruitment*" (recruited for intervention versus not recruited for intervention) as covariates, in order to rule out any confounding effects of early intervention. Three tests were used: (i) main effect of *recruitment* to account for any differences due to sampling of the groups; (ii) *intervention* x condition interaction to test for any effect intervention might have had on face/gaze ERPs (iii) *treatment* x condition interaction to account for any moderating effect treatment has on face/gaze processing ERPs.

FACE/NOISE CONTRAST

A significant interaction between condition and group emerged on N290 latency (F (2,169)=3.95, p=.021). For this finding, there was no significant effect of recruitment into intervention (F(1,169)=0.73, p=.396) and no significant interaction between outcome and intervention (F(1,169)=1.20, p=.275), nor treatment (F(1,160)=1.71, p=.193).

DYNAMIC GAZE CONTRAST

There was a significant condition x outcome interaction on P1 latency (F (2, 211) = 5.06, p=.007), but no main effect of recruitment into intervention (F(1,211)=0.19, p=.664), nor interaction between condition and intervention (F(1,211)=0.30, p=.582), and treatment (F(1,211)=0.02, p=.885).

For the significant interaction between condition and group on P400 latency (F (2, 206)= 3.61, p=.029), there was also no main effect of recruitment (F(1,206)=0.38, p=.539), nor interaction between outcome and intervention (F(1,206)=1.54, p=.213) and treatment (F(1,206)=2.48, p=.117).

There was a significant condition x outco6e interaction on P400 amplitude (F (2,210)=3.64, p=.028, but no significant effect of recruitment (F(1,210)=0.01, p=.923) nor interaction

between outcome and intervention (F(1,210)=0.01, p=.941) and treatment (F(1,210)=0.03, p=.866).

2) COHORT 2-ONLY ANALYSIS

FACE/NOISE CONTRAST

A significant condition x outcome interaction emerged on N290 latency (F (2, 96) = 4.80, p=.010). The HR-ASD group showed a diminished effect of face versus noise compared to the LR (p=.003, d=1.12) and HR-no ASD (p=.016, d=0.71) groups, with no difference between LR and HR-no ASD (p=.162, d=0.38). There were no other significant interactions (all ps>.05).

DYNAMIC GAZE CONTRAST

There was a significant condition x outcome interaction on P400 latency (F (2,114)=3.55, p=.03), whereby the LR and HR-no ASD groups showed longer latency to gaze shifting away versus towards, compared to HR-ASD (LR: p=.010, d=0.92; HR-no ASD: p=.045, d=0.53). There was no significant difference between LR and HR-no ASD (p=.187, d=0.35). No other interactions were significant (all ps>.05).

STATIC GAZE CONTRAST

There was a significant condition x outcome interaction on P400 latency (F (2,64)=3.55, p=.035), whereby the LR group showed a longer latency to direct static gaze compared to averted gaze, compared to the HR=no ASD (p=.021, d=0.88) and the HR-ASD groups (p=.015, d=1.40). There was no significant difference between the HR groups (p=.042, d=0.31). There were no other significant interactions for the static gaze contrast (all ps>.05)

3) CONFOUNDING VARIABLES IN THE INDIVIDUAL-LEVEL ANALYSIS

Gender was included in the initial pool of features for selection and was always selected by the best performing classifiers (AUC>85%) in repeated evolution (frequency=100%). To investigate whether the selected features for prediction of ASD outcome were actually predicting outcome and not gender, we used the optimal set of features selected by the genetic algorithm (see Table 1 in the main text) to classify males from females. Gender was however excluded from the optimal set to allow prediction of gender. Results are shown in Table S3. Given the poor classification performance, we could exclude that our optimization algorithm was affected by gender bias and that classifiers were predicting gender.

Age was included in the initial pool of features for selection and its incidence among the peaks of repeated evolution was considered as an indicator of relevance for prediction of ASD outcome. A frequency of less than 10% indicated that age was not discriminative of ASD outcome and was not included in the main classification analysis.

To check for the confounding effect of MSEL non-verbal T-scores (NVT-scores) at 36 months (i.e. fine motor and visual reception scores) on classification of ASD outcome, we used the optimal set of features selected for classification of HR-ASD to classify HR siblings with NVT-scores more than 1.5 standard deviations below average (<35) from HR siblings with average or above average scores at 36 months. Results (see Table S3) showed that prediction of NVT-scores was not significantly different from chance level, thus we can exclude that our classifiers were predicting non-verbal skills rather than clinical outcome. Of note, AUC was at trend level significance and accuracy was significantly different from random prediction for visual reception scores; however, performance of the classifier for prediction of visual receptive skills was significantly lower than for ASD outcome (Wilcoxon p<.001; respectively z=-38.7 for AUC and z=-44.7 for accuracy).

4) MOST RELEVANT FEATURES AMONG BEHAVIOURAL AND ERP DATA FOR PREDICTION OF ASD

Among the 145 HR infants used in the main analysis of ERP data, a subsample of 140 HR infants was selected based on having behavioural and ERP data available at 8 months (see Table S7 for demographics and clinical characteristics of the sample).

TABLE S7: Demographics and clinical characteristics.

	HR-ASD (n=30)	HR-no ASD (n=110)
Gender	**n**	**n**
Males	23	48
Females	7	62
Cognitive development	Mean (standard deviation)	Mean (standard deviation)
Gross motor skills	42,82 (10,29)	45,75 (11,24)
Fine motor skills	47,80 (11,77)	54,52 (12,81)
Visual receptive skills	50,46 (9,16)	55,44 (11,79)
Receptive language skills	42,80 (12,02)	48,09 (9,85)
Expressive language skills	49,82 (11,64)	50,69 (10,56)
Adaptive behavior	Mean (standard deviation)	Mean (standard deviation)
Communication skills	89,77 (15,46)	94,75 (15,63)
Daily living skills	93,17 (15,16)	100,60 (12,82)
Motor skills	84,17 (15,99)	86,53 (16,70)
Socialization skills	96,87 (15,91)	99,55 (11,79)
ASD symptoms severity	Mean (standard deviation)	Mean (standard deviation)
Total score	10,90 (5,43)	9,05 (4,75)

Behavioural data at 8 months included:

- Measures of cognitive development from the Mullen Scales of Early Learning (MSEL,[1]), a standardized developmental measure assessing cognitive functioning in 5 scales (T-scores: mean=50; standard deviation, SD=10): gross motor (GM), visual reception (VR), fine motor (FM), receptive (RL) and expressive language skills (EL).
- Measures of adaptive behavior from the Vineland Adaptive Behavior Scales (VABS-II,[2]), a semi-structured parent-report questionnaire assessing personal and social functioning in 4 different domains (Standard scores: mean=100; SD=15): Communication (Comm), Daily Living Skills (DL), Socialization (Soc) and Motor Abilities (Mot).
- ASD symptoms severity as measured by the total score of the Autism Observation Scale for Infants (AOSI). We used a 19-item version of this semi-structured observational assessment[138] to detect putative behavioral signs of ASD.

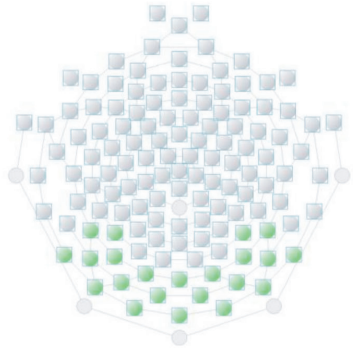
We performed feature selection using a genetic algorithm to extract information about the most relevant features for prediction of ASD. A total of 65 variables including gender, behavioural measures, averaged ERP measures in response to each condition, and differential ERP responses for each contrast were used as features for the classifier analysis (see Table S8 for the list of features). This analysis investigated which features among behavioural and neural measures at 8 months are the most relevant for prediction of ASD outcome at 36 months among HR siblings. Each feature was standardized, and we used a genetic algorithm[183] to select the features for the classifier. We chose the Area Under the Curve (AUC) from a 10-fold cross-validated SVM classifier built on the set of features under evaluation as the target value for fitness. Population size [n=100] and number of generations [n=200] were selected by an experienced researcher (KJ). The number of features for the classifier, was selected based on the AUC level reached during the evolutionary process, and stability of the process assessed through visual inspection. Once selected the number of features (n=16), the evolutionary process was repeated n=100 times to investigate the variability in the feature space.

The candidate solutions with highest AUC (higher than 85%) were selected and used for frequency analysis on the selected features, as the incidence frequency of each feature provides an estimate of the relevance of each feature for the specific classification problem. The features with highest incidence (higher than 80%) were selected as the most relevant for prediction of ASD and included: gender, amplitude of the P400 component in response to faces with averted gaze (FA P400 amplitude), latency of the P400 component in response to gaze shifts directed towards the infant compared to shifts directed away (SD-SA P400 latency), latency of the N290 component in response to faces with direct compared to averted gaze (FD-FA N290 latency), amplitude of the P1 component in response to visual noise compared to faces (N-F P1 amplitude), and latency of the N290 component in response to visual noise compared to faces (N-F N290 latency). None of the behavioural features was selected among the most relevant features for prediction of ASD at 8 months. Thus, results suggest that there is more predictive power retained by neural measures compared to behavioural measures for prediction of later ASD outcome at such an early age.

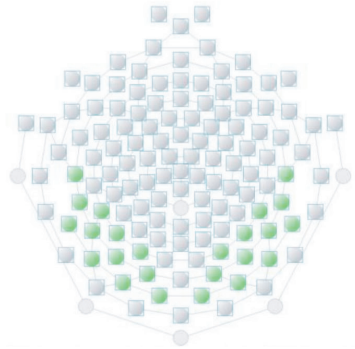
TABLE S8. Behavioural and ERP features for the classification problem.

SD p1 amplitude	SD p1 latency	FD p1 amplitude	FD p1 latency	N p1 amplitude	N p1 latency
SA p1 amplitude	SA p1 latency	FA p1 amplitude	FA p1 latency	F p1 amplitude	F p1 latency
SD n290 amplitude	SD n290 latency	FD n290 amplitude	FD n290 latency	N n290 amplitude	N n290 latency
SA n290 amplitude	SA n290 latency	FA n290 amplitude	FA n290 latency	F n290 amplitude	F n290 latency
SD p4 amplitude	SD p4 latency	FD p4 amplitude	FD p4 latency	N p4 amplitude	N p4 latency
SA p4 amplitude	SA p4 latency	FA p4 amplitude	FA p4 latency	F p4 amplitude	F p4 latency
SD-SA p1 amplitude	SD-SA p1 latency	FD-FA p1 amplitude	FD-FA p1 latency	N-F p1 amplitude	N-F p1 latency
SD-SA n290 amplitude	SD-SA n290 latency	FD-FA n290 amplitude	FD-FA n290 latency	N-F n290 amplitude	N-F n290 latency
SD-SA p4 amplitude	SD-SA p4 latency	FD-FA p4 amplitude	FD-FA p4 latency	N-F p4 amplitude	N-F p4 latency
Gender	Gross motor skills	Fine motor skills	Visual receptive skills	Receptive language	Expressive language
Communication skills	Daily living skills	Motor skills	Socialization skills	Symptoms severity	

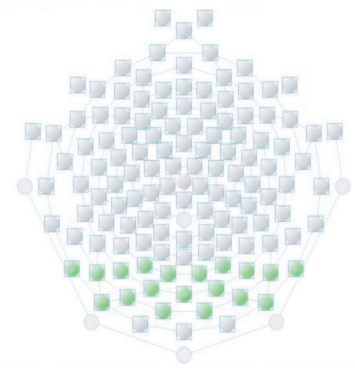
This table shows the list of 10 behavioural features + 54 ERP features + gender (n=65 in total) used as input features for the genetic algorithm to select features for classification. *Abbreviations:* SD = gaze shift directed towards the infant; SA = gaze shift directed away from the infant; FD = static face with direct gaze towards the infant; FA = static face with gaze directed away from the infant; F = static face (average direct and averted gaze); N = visual noise.



Static gaze contrast montage



Dynamic gaze contrast



Face/noise contrast

FIGURE S1: Selected channel montages based on Elsabbagh et al. (2012) and corroborated with visual inspection of grand averages.

PART II

Parsing heterogeneity in autism spectrum disorder

5

Latent trajectories of adaptive behaviour in infants at high and low familial risk for autism spectrum disorder

Published as: "Latent trajectories of adaptive behaviour in infants at high and low familial risk for autism spectrum disorder", Mol Autism, 2019.

Bussu, G.,
Jones, E.J.H.,
Charman, T.,
Johnson, M.H.,
Buitelaar, J.K. & The BASIS Team

ABSTRACT

BACKGROUND

Autism Spectrum Disorder (ASD) is characterized by persisting difficulties in everyday functioning. Adaptive behaviour is heterogeneous across individuals with ASD, and it is not clear to what extent early development of adaptive behaviour relates to ASD outcome in toddlerhood. This study aims to identify subgroups of infants based on early development of adaptive skills and investigate their association with later ASD outcome.

METHODS

Adaptive behaviour was assessed on infants at high (n=166) and low (n=74) familial risk for ASD between 8 and 36 months using the Vineland Adaptive Behavior Scales (VABS-II). The four domains of VABS-II were modelled in parallel using growth mixture modelling to identify distinct classes of infants based on adaptive behaviour. Then, we associated class membership with clinical outcome and ASD symptoms at 36 months, and longitudinal measures of cognitive development.

RESULTS

We observed three classes characterised by: decreasing trajectories below age-appropriate norms (8.3%); stable trajectories around age-appropriate norms (73.8%); increasing trajectories reaching average scores by age 2 (17.9%). Infants with declining adaptive behaviour had a higher risk [odd ratio, OR=4.40 (confidence interval, CI: 1.90; 12.98)] for ASD and higher parent-reported symptoms in the social, communication and repetitive behaviour domains at 36 months. Furthermore, there was a discrepancy between adaptive and cognitive functioning as the class with improving adaptive skills showed stable cognitive development around average scores.

CONCLUSIONS

Findings confirm the heterogeneity of trajectories of adaptive functioning in infancy, with a higher risk for ASD in toddlerhood linked to a plateau in the development of adaptive functioning after the first year of life.

INTRODUCTION

Autism Spectrum Disorder (ASD) is a set of heterogeneous developmental disorders characterised by difficulties in the social-communication domain, restricted and repetitive patterns of behaviours and interests, and sensory anomalies[8]. A diagnosis of ASD according to the DSM-5 criteria further requires that symptoms cause clinically significant impairments in everyday functioning. The resulting long-term outcome is mixed but many children diagnosed with ASD will have a sub-optimal quality of life in adulthood[188-190], with persisting impairments in everyday functioning[18]. Adaptive behaviour reflects the ability of an individual to function independently in everyday situations and reflects how an individual responds to environmental demands translating capacities into everyday competencies. Evidence suggests that adaptive behaviour in ASD is heterogeneous across individuals and even within the same individual over time[191]. A useful approach to explore this heterogeneity is growth mixture modelling (GMM). GMM is a data-driven approach which allows one to identify distinct mixtures of trajectories within population defining classes of individual growth curves[125]. Previous studies have used GMM to parse samples based on the heterogeneity in development of adaptive behaviour, mainly focusing on pre-school and school-aged children with ASD[191-195]. These studies analysed approximately 100 children with ASD to derive classes of developmental trajectories on the Adaptive Behaviour Composite (ABC) score[192, 193] and the Daily Living score[194, 195] of the Vineland Adaptive Behaviour Scale (VABS-II[196]). Only one study[191] used a larger sample to identify classes in ABC scores (n=421). Results mainly showed one class with low functioning and decreasing scores, another class with moderate functioning and stable scores over time, and one final class with higher functioning and substantial increase of scores over time. The partial replication of classes across these studies suggests that developmental trajectories of adaptive skills may actually serve to define subgroups in the ASD phenotype.

The work reviewed above focused on children from approximately 3 years of age who already had a diagnosis of ASD. Thus, it could not capture the different developmental trajectories preceding an ASD diagnosis. The investigation of early development of adaptive skills is critical to predict later functional outcome and then enable early targeted intervention. Early access to personalized interventions can, in turn, be crucial to improve later functioning in different environments of everyday life through learning of adaptive skills[197]. This stresses the importance of investigating not only the presence and development of ASD symptoms, which can even be masked by learned strategies to cope with environmental demands[8], but also the different developmental pathways of adaptive behaviour that infants can follow early in life.

However, adaptive behaviour has been rarely used so far as an outcome measure in prospective research. A recent study used mixed-models to examine developmental trajectories of cognitive and adaptive functioning between 7 months and 7 years of age, measured respectively by the Mullen Scales of Early Learning (MSEL) and the VABS-II, in infants at familial high-risk (HR) and low-risk (LR) for ASD[59]. Results showed that HR siblings who met criteria for ASD at age 7 had increasing difficulties in adaptive behaviour over time compared to LR controls, while HR siblings without later ASD outcome did not differ in adaptive behaviour from LR siblings. These findings extend previous work on the same dataset[55] and independent work from Estes and colleagues[35] on trajectories of adaptive behaviour between 8 and 36 months. These studies showed decreased adaptive functioning by 12 months in high-risk siblings developing ASD compared to non-ASD siblings and low-risk controls. There is a partial overlap in data between the study from Salomone et al[59] and the present study; however, analytic methods and research aims were different. In fact, although these findings improve our understanding of the different developmental profiles of adaptive functioning in infancy and toddlerhood, analyses were based on clinical outcome groups and might not have captured the variation in phenotypes within clinical categories. Only one recent study has investigated latent trajectories of adaptive functioning in a high-risk cohort[198]. This study analysed 566 infants between 12 and 36 months to derive classes of developmental trajectories on the ABC score. Results showed one class with average scores at 12 months and a declining trajectory, one class with a slightly declining trajectory, and one class with higher scores and a stable trajectory. The current study aimed to identify distinct classes of infants based on early development of adaptive functioning in HR siblings and LR controls. We used a prospective analysis approach to discover structure in data, independently from clinical categories. Specifically, we used parallel process GMM to simultaneously examine communication, daily living, motor and social domains of adaptive behaviour, and observe strengths and impairments in the different areas of everyday functioning. Then, we characterised the identified classes in terms of ASD outcome and symptoms at 36 months and concurrent trajectories of cognitive development. Previous studies have shown a discrepancy between adaptive functioning, cognitive abilities and ASD symptoms in older children with ASD[191, 199-201]. These findings suggest that neither normative cognitive development nor low levels of ASD symptoms are protective factors against poor adaptive functioning. Our post-hoc association of class membership with cognitive development and symptom severity allowed us to investigate the relationship between these three areas of functioning in early development. Overall, the exploratory investigation of inter-individual heterogeneity with such an unsupervised approach provides better insight into the variety of paths leading to different functional outcomes within ASD and typical development.

METHODS

PARTICIPANTS AND PROCEDURE

Data were collected from 247 infants from the British Autism Study of Infant Siblings[202], across two phases of the study based on time of recruitment. Infants were considered at high (n=170) and low (n=77) familial risk for ASD based on having or not an older biological sibling with ASD. 54 high-risk and 50 low-risk infants participated to Phase 1[87], while an independent cohort of 116 HR and 27 LR participated to Phase 2. LR controls were full-term infants recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. At the time of enrolment, none of the infants had been diagnosed with any developmental condition. Infants were followed longitudinally on four visits: 8 months [mean=8.1; standard deviation, SD=1.2], 14 months [mean=14.5, SD=1.3], 24 months [mean=25.0, SD=1.8] and 36 months [mean=38.8, SD=3.0]. To allow testing for quadratic growth, the final sample included infants with data available from at least 3 visits, leading to a final sample of 240 infants (74 LR and 166 HR). Study researchers were aware of infants' risk status but were blind to clinical outcome.

MEASURES

Adaptive functioning

The Vineland Adaptive Behavior Scale (VABS-II[196]) is a semi-structured parent-report questionnaire (at 8 and 14 months) or parent interview (at 24 and 36 months) assessing infant's adaptive behaviour in everyday settings. The items address personal and social functioning in 4 different domains (standard scores; mean = 100, SD = 15): Communication (Comm), Daily Living Skills (DL), Socialization (Soc) and Motor Abilities (Mot). Standard scores from the 4 domains between 8 and 36 months were included in our main analysis to identify homogeneous classes of infants.

Developmental skills

Verbal and non-verbal cognitive development was measured at each visit by the Mullen Scales of Early Learning (MSEL[203]), a standardized developmental measure used to assess cognitive functioning between birth and 68 months. Scores are obtained in 5 domains: gross motor (GM), visual reception (VR), fine motor (FM), receptive (RL) and expressive language skills (EL). The Mullen Scale provides normative scores for each scale (T-scores: mean = 50, SD = 10) and a single composite score representing general intelligence (Early Learning Composite, ELC standard score: mean = 100, SD = 15). ELC-scores between 8 and 36 months were included in our analyses to characterise the developmental level of the identified classes.

Early ASD symptoms

The Autism Diagnostic Observation Schedule (ADOS[204]), a standardised diagnostic instrument, the Autism Diagnostic Interview – Revised (ADI-R[7]), a structured parent interview, and the Social Communication Questionnaire (SCQ[205]), a screening tool for ASD, were administered at 36 months to assess autism symptoms. Of note, the ADI-R was not administered to LR infants from Phase 1 (n=47) and missing values were imputed through expectation maximization on SPSS[206].

To evaluate the end level of symptom severity of the identified classes, we included in our analysis: the ADOS Calibrated Severity Score (CSS) obtained from the raw total scores (CSS-Tot), and Social Affect (CSS-SA) and Restricted and Repetitive Behaviors (CSS-RRB) domains; the ADI-R domain scores for the Social (ADI-Soc), Communication (ADI-Comm) and Repetitive Behaviours and Interests domains (ADI-RBI); and the SCQ total score (SCQ-Tot).

CIINICAL OUTCOME

The LR group was based on having an older full sibling with typical development. LR infants received no formal clinical diagnoses, but none of them had a community clinical ASD diagnosis at 36 months. In particular, no ADI-R was administered to LR in Phase 1, who did not receive an outcome evaluation. In Phase 2, LR infants were administered the ADOS and ADI-R and received an outcome evaluation at 36 months, but none of them raised any concern for ASD or atypical development. HR siblings received a clinical outcome evaluation at 36 months and were subsequently grouped into siblings with ASD (HR-ASD); with atypical (non-ASD) development (HR-Atypical); and with typical development (HR-Typical).

Expert clinical researchers reviewed all available information at 24 months and 36 months and assigned clinical consensus best estimate diagnosis of ASD (HR-ASD) according to ICD-10[9] or DSM-5 criteria depending on the recruitment phase[8]. Diagnoses were reviewed for differences in categorisation and considered to be similar. Among high-risk infants who did not meet criteria for ASD, a subgroup of siblings was classified as 'atypical' (HR-Atypical) based on: ADOS and/or ADI-R above ASD threshold, and/or MSEL more than 1.5 standard deviations below average on visual reception and/or receptive language and/or expressive language and/or early learning composite (n=15) scores. Finally, siblings who did not meet criteria for ASD or atypical development were classified as HR-Typical.

DATA ANALYSIS: AN OVERVIEW

We used a three-step approach to identify latent classes of adaptive behaviour and profile them through associations with external variables. First, the four domains of the Vineland were modelled in parallel through growth mixture modelling to identify latent class trajectories of adaptive behaviour on 4 time-points between 8 and 36 months. Second, infants were assigned to latent classes based on posterior probabilities of class membership. Third, the identified classes were characterised in terms of clinical outcome and symptom severity at 36 months, and longitudinal cognitive development.

IDENTIFICATION OF LATENT CLASS TRAJECTORIES

We chose growth mixture modelling to identify distinct mixtures of trajectories within population. As opposed to other methods such as latent class growth curve modelling[123], which assumes a homogeneous pattern of behaviour within class, growth mixture modelling[125] allowed us to capture the complexity of adaptive behaviour in developmental variation across individuals.

We investigated the pattern of missing data for the four domains of adaptive behaviour by testing its association with gender and clinical outcome at 36 months. Differences in gender were not significant, while the proportion of missing data at 24 months was significantly dependent on clinical outcome at 36 months ($\chi^2(3)=8.23$, p=0.04), with HR-Atypical having most missing data. However, differences in outcome were not significant at other time-points, providing reasonable evidence for a pattern of data missing at random. Thus, individuals with missing data were included in the analysis, allowing us to use all available information. In fact, individual trajectories of adaptive behaviour were modelled on data available at an individual level.

Real age was included as a fixed effect while random effects on intercept and slope were modelled on an individual level. Multiple models were tested based on the polynomial degree of the growth curve, the variance/covariance matrix and the number of classes. Models were run with 1 to 6 classes and each class number was run separately 50 times to control for local maxima. The best model was determined in terms of data fitting and parsimony based on having lower values of Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and negative log-likelihood, and higher average class posterior probability[207]. Analyses were performed using the *multlcmm* function from the *lcmm* package in R[208].

The classes derived from parallel process growth mixture modelling were subsequently compared on adaptive behaviour over time through hierarchical mixture modelling[139]. A quadratic mixed model was tested with VABS-II domain scores as outcome variables and real age and class membership as fixed factors, while gender was included as a covariate and random effects on intercept and slope were modelled on an individual level. We investigated the main effects of *class*, *age*, *age*^{2} and their interaction effects using Wald

tests with Satterthwaite approximation for degrees of freedom. Post-hoc Tukey's tests for multiple comparisons were performed for class comparisons and simple main effects analysis. Analyses were implemented using the *lme4* software package on R[140].

CHARACTERIZATION OF LATENT CLASSES

Classes in adaptive behaviour, as derived from parallel process growth mixture modelling, were further characterised by examining the association of class membership with independent outcome variables. First, we examined the association with ASD symptom severity at 36 months, as measured by the CSS-Tot, CSS-SA, CSS-RRB, ADI-Comm, ADI-Soc, ADI-RBI and SCQ-Tot scores, through an analysis of variance. For significant differences, classes were compared through post-hoc Tukey's tests for multiple comparisons.

Then, we examined the association of class membership with trajectories of cognitive development, as measured by the MSEL ELC score between 8 and 36 months, through hierarchical mixture modelling[139]. Models were built using the *lme4* software package on R[140], with MSEL ELC scores as outcome variables, real age and class membership as fixed factors, and gender as a covariate, while random effects on intercept and slope were modelled on an individual level. We compared linear and quadratic models on age to select the best fit based on chi-squared tests on the log-likelihood values.

RESULTS

Among HR siblings at 36 months, 34/166 [20.5%] siblings were categorised as HR-ASD, 44/166 [26.5%] as HR-Atypical, and 87/166 [52.4%] as HR-Typical. Among HR-Atypical, 32/87 and 6/87 siblings respectively had ADOS and ADI-R scores above ASD threshold, 9/87 siblings had MSEL more than 1.5 standard deviations below average on visual reception, 14/87 on receptive language, 9/87 on expressive language, and 15/87 on early learning composite scores. Finally, 1/166 infant sibling did not receive a clinical outcome evaluation but was included in our trajectory analysis having complete data on adaptive behaviour. Descriptive statistics for the entire sample and the classes identified are shown in Table 1, while descriptive statistics by risk group are reported in Table S1 (see Additional file 1). Three classes of quadratic trajectories provided the best fit to the data, with BIC=28523.12, AIC=28397.82 and average posterior probability of 87%. Metrics of model fitting are reported in Table S2 (see Additional file 1). The identified trajectories of adaptive behaviour are shown in Figure 1. Modelling the corresponding trajectories in ELC scores (Figure 2), the quadratic model was the best fit for the data ($\chi^2(6)=26.2$, $p<0.001$).

TABLE 1: Descriptive statistics. This table shows descriptive statistics for the entire sample and for the different trajectory classes of adaptive behaviour.

	tot		c1 decreasing adaptive behaviour		c2 average/stable adaptive behaviour		c3 recovering adaptive behaviour	
	mean	sd	mean	sd	mean	sd	mean	sd
Age								
8m	8.1	1.2	7.9	1.5	7.9	1.1	8.7	1.3
14m	14.4	1.3	14.1	1.4	14.3	1.2	15.1	1.4
24m	25.0	1.8	25.7	2.7	25.0	1.8	25.0	1.5
36m	38.8	2.9	40.2	3.6	38.6	3.0	38.9	2.2
VABS								
Comm 8m	96.1	15.9	108.2	14.7	99.2	12.6	77.0	14.2
Comm 14m	96.7	13.3	91.1	15.9	98.0	13.1	93.4	11.8
Comm 24m	103.7	13.0	88.7	20.1	104.7	11.3	106.6	11.2
Comm 36m	101.1	14.1	86.8	19.8	102.2	12.9	102.9	12.6
DL 8m	100.1	13.7	106.7	18.4	101.3	12.3	91.9	13.8
DL 14m	95.2	13.1	90.6	17.8	96.5	13.0	91.7	10.0
DL 24m	105.7	12.9	94.4	20.0	106.7	11.9	106.7	10.1
DL 36m	103.2	13.0	87.9	20.9	104.1	11.5	106.2	9.3
Mot 8m	89.8	16.3	98.4	24.2	91.8	14.5	76.9	12.7

TABLE 1 CONTINUED.

	tot		c1 decreasing adaptive behaviour		c2 average/stable adaptive behaviour		c3 recovering adaptive behaviour	
	mean	sd	mean	sd	mean	sd	mean	sd
VABS								
Mot 14m	100.3	12.8	99.1	14.9	101.9	12.5	94.3	11.4
Mot 24m	100.1	10.9	90.2	12.0	101.2	10.4	100.4	10.1
Mot 36m	93.8	12.4	84.6	14.7	94.7	11.9	94.1	11.7
Soc 8m	100.1	12.8	109.6	19.7	101.4	10.6	90.0	11.7
Soc 14m	97.7	11.7	97.5	13.8	98.7	11.8	93.6	9.7
Soc 24m	101.0	11.6	89.1	18.0	101.5	10.6	104.1	8.7
Soc 36m	97.8	12.9	85.6	17.9	98.3	12.1	101.2	10.5
MSEL								
ELC 8m	104.2	15.0	106.2	18.2	104.9	14.7	100.4	14.6
ELC 14m	98.5	16.0	86.2	19.3	100.6	15.7	95.6	12.6
ELC 24m	104.7	19.9	81.9	25.8	106.8	18.0	105.7	19.1
ELC 36m	107.7	22.7	89.6	26.7	108.6	22.7	111.9	17.4
ADOS at 36m[1]								
CSS-Tot	2.95	2.40	3.84	3.00	2.94	2.43	2.60	1.90
CSS-SA	3.40	2.51	4.00	2.89	3.43	2.55	3.02	2.14
CSS-RRB	4.29	2.61	5.42	2.73	4.14	2.58	4.42	2.62
ADI-R at 36m[2]								
ADI-Comm[3]	3.25	4.17	7.00	5.75	2.96	3.85	2.67	3.81
ADI-Soc[3]	3.08	4.31	7.55	7.76	2.66	3.68	2.67	3.35
ADI-RBI[4]	1.15	1.98	2.40	2.58	1.12	1.93	0.70	1.63
SCQ at 36m[2]								
SCQ-Tot[3]	5.32	6.11	11.1	8.83	5.02	5.64	3.91	5.03
Clinical outcome at 36m[5]	n (%)		n (%)		n[6] (%)		n (%)	
LR	74 (31%)		4 (20%)		56 (32%)		14 (32%)	
HR-Typ	87 (36%)		6 (30%)		63 (36%)		18 (42%)	
HR-Atyp	44 (18%)		2 (10%)		34 (19%)		8 (19%)	
HR-ASD	34 (14%)		8 (40%)		23 (13%)		3 (7%)	
Gender								
Female	122 (51%)		7 (35%)		92 (52%)		23 (53%)	
Male	118 (49%)		13 (65%)		85 (48%)		20 (47%)	

Abbreviations. tot = entire sample; c1-c3 = classes in trajectories of adaptive behaviour[7]; VABS = Vineland Adaptive Behavior Scales; Comm = communication score; DL = daily living score; Mot = motor score; Soc = socialization score; MSEL = Mullen Scales of Early Learning; ELC = early learning composite score; ADOS = Autism Diagnostic Observation Schedule; CSS = calibrated severity score; ADI-R = Autism Diagnostic Interview-Revised; ADI-Comm = Communication domain score (ADI-R); ADI-Soc = Social domain score (ADI-R); ADI-RBI = Restricted Behaviors and Interests domain score (ADI-R); SCQ = Social Communication Questionnaire; SCQ-Tot = Total score (SCQ); LR = low-risk controls; HR = high-risk siblings; HR-Typ = typically developing siblings; HR-Atyp = atypically developing siblings (no ASD); HR-ASD = siblings with ASD.

[1] Data were available for a subsample of n=235 infants.

[2] Data were available for a subsample of n=239 infants.

[3] Significant difference per class with p<0.001.

[4] Significant difference per class with p<0.05.

[5] Clinical outcome vs class membership: $\chi^2(6)=13.39$, p=0.037.

[6] One infant in this class did not receive a clinical outcome evaluation at 36 months.

[7] c1 = decreasing adaptive behaviour; c2 = average/stable adaptive behaviour; c3 = recovering adaptive behaviour.

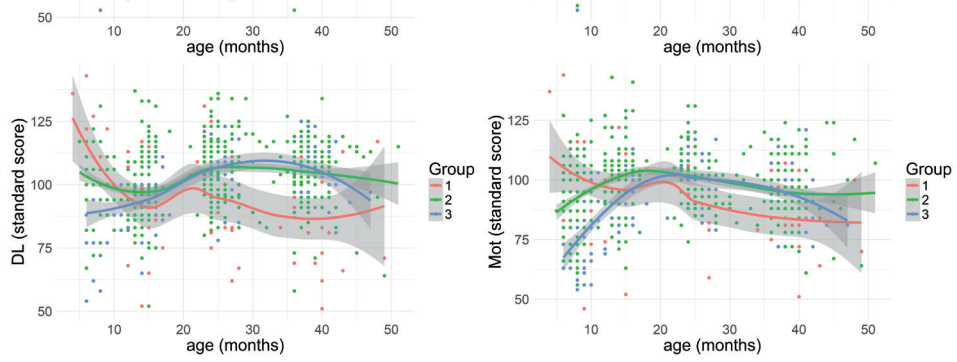


FIGURE 1: latent trajectories in adaptive behaviour. This figure illustrates the 3 classes identified in longitudinal adaptive behaviour. Points show individual scores while classes were computed through the *loess* function in R for visualization purposes. Class 1 shows a decreasing trajectory in all scales; class 2 shows a stable trajectory at an average level of adaptive behaviour; class 3 shows recovering (improving) trajectories, starting from low scores in all scales and reaching an average level from around 20 months onwards.

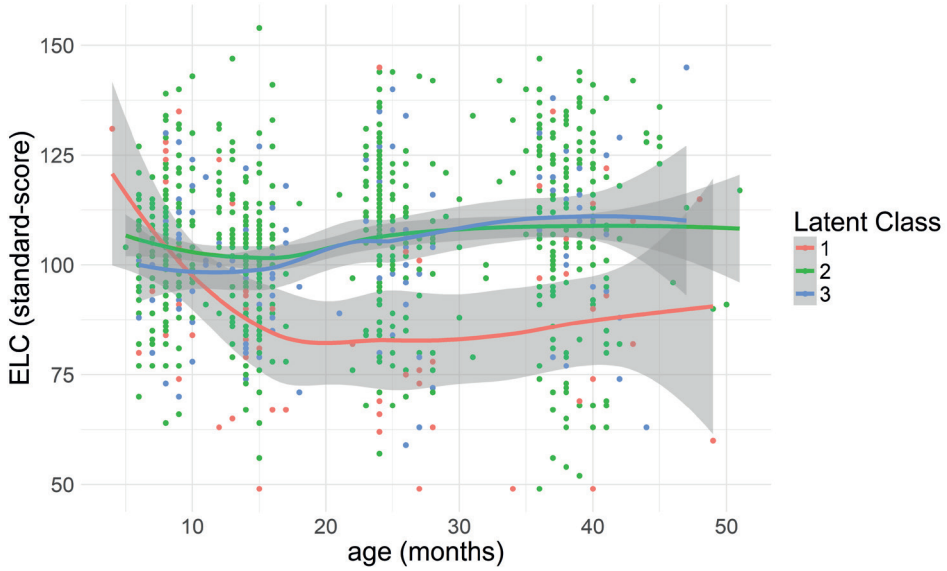


FIGURE 2: Developmental trajectories in cognitive level. This figure illustrates the developmental trajectories of cognitive level, as measured by the ELC score from the Mullen Scales of Early Learning, for the 3 classes identified in longitudinal adaptive behaviour. Points show individual scores while classes were computed through the *loess* function in R for visualization purposes. Class 1 shows a decreasing trajectory in cognitive level; class 2 and 3 show stable trajectories at an average level of cognitive development.

Class 1 [n_1=20 (8.3%)] shows decreasing trajectories in all domains of adaptive behaviour (fixed effect of age in the common longitudinal model: β=-0.27; standard error, SE=0.08; p=0.001) starting with above average age standardized scores in communication, daily living and socialization skills, and average scores in motor skills in the first year of life. This is the only class with unbalanced gender, being more males (65%) than females (35%). Results from linear mixed modelling show that infants in Class 1 had significantly decreasing standardized scores in communication, daily living and socialization domains between 8 and 24 months (p<0.001), and in socialization between 24 and 36 months (p=0.04), while motor scores decreased significantly only between 14 and 24 months (p=0.02). Infants in this class also show a similar trajectory in cognitive development over time, with decreasing ELC scores before age 2 (p<0.001), starting from above average scores in the first year of life.

Class 2 [n_2=177 (73.8%)] shows stable trajectories around average scores in all domains of adaptive behaviour (fixed effect of age in the common longitudinal model: β=0.11; SE=0.03; p<0.001). Gender was balanced in this class, with 48% of males and 52% of females. Results from linear mixed modelling show that infants had significantly increasing age standardized scores between 8 and 14 months in communication (p=0.004) and daily living domains (p=0.04), increasing scores between 14 and 24 months in communication, daily living and motor domains (p<0.001), and decreasing motor scores between 24 and 36 months (p<0.001). Infants in this class show a stable trajectory around average scores in cognitive development before age 2 and increasing ELC scores afterwards(p=0.005).

Class 3 [n_3=43 (17.9%)] shows recovering (improving) trajectories (fixed effect of age in the common longitudinal model: β=0.43; SE=0.10; p<0.001) starting from low scores and reaching a stable average level in all domains by age 2. Gender was balanced in this class, with 47% of males and 53% of females. Results from linear mixed modelling show that infants had significantly increasing age standardized scores between 8 and 24 months in all domains of adaptive behaviour (p<0.001). Infants in this class also show a stable trajectory around average scores in cognitive development before age 2 with increasing ELC scores afterwards (p=0.007).

Mixed models on trajectories of adaptive behaviour show that Class 3 had significantly lower scores than Classes 1 and 2 in all domains at 8 months (p<0.001), while it had only significantly lower scores than Class 2 in communication and motor scores at 14 months (p<0.001). Class 1 had significantly lower scores than Classes 2 and 3 in communication (ps<0.001), daily living (respectively p=0.002 and p=0.006), motor (respectively p=0.004 and p<0.05), and socialization scores (respectively p=0.006 and p=0.002) at 24 months. Finally, Class 1 had significantly lower scores than Classes 1 and 2 in communication (ps<0.001), daily living (ps<0.001), motor (respectively p=0.002 and p=0.007) and socialization scores (ps<0.001) at 36 months (see Figure 1).

RELATIONSHIP BETWEEN CLASSES AND COGNITIVE DEVELOPMENT.

Infants in Class 1 had significantly lower scores on cognitive development, measured by the MSEL, than those in Class 2 at 14 months (p=0.01), in Class 2 (p<0.001) and 3 (p=0.002) at 24 months, and in Class 2 and 3 at 36 months (ps<0.001); while Classes 2 and 3 identified from longitudinal development of adaptive behaviour did not differ in cognitive development (see Figure 2). Thus, there was a split between adaptive skills and cognitive skills for the improving class.

RELATIONSHIP BETWEEN CLASSES AND CLINICAL OUTCOME AT 36 MONTHS.

TABLE 2: Class differences in symptom severity at 36 months. This table shows detailed statistics from the ANOVA investigated differences between classes in symptoms severity at 36 months as measured by the ADOS, ADI-R and SCQ. When differences were significant, Post-hoc Tukey's tests for multiple comparisons were performed for paired comparisons of classes. Differences were considered significant for $p<0.05$ (marked as *).

	ANOVA			Post-hoc Tukey's Tests					
				Class 2 vs. Class 1		Class 3 vs. Class 1		Class 3 vs. Class 2	
	F	dof	p	t	p	t	p	t	p
ADOS at 36m[1]									
CSS-Tot	1.77	2/232	0.17	-	-	-	-	-	-
CSS-SA	1.04	2/232	0.35	-	-	-	-	-	-
CSS-RRB	2.15	2/232	0.12	-	-	-	-	-	-
ADI-R at 36m[2]									
ADI-Comm[3]	9.56	2/236	<0.001*	-4.25	<0.001*	-3.97	<0.001*	-0.42	0.91
ADI-Soc[3]	13.0	2/236	<0.001*	-5.04	<0.00*	-4.39	<0.001*	0.01	1.00
ADI-RBI[4]	5.32	2/236	0.005*	-2.79	0.015*	-3.23	0.004*	-1.28	0.40
SCQ at 36m[2]									
SCQ-Tot[3]	11.2	2/236	<0.001*	-4.40	<0.001*	-4.53	<0.001*	0.50	0.50

Abbreviations. ANOVA = ANalysis Of VAriance; Class 1 = decreasing adaptive behaviour; Class 2 = average/stable adaptive behaviour; Class 3 = recovering adaptive behavior; dof = degrees of freedom; ADOS = Autism Diagnostic Observation Schedule; CSS = calibrated severity score; ADI-R = Autism Diagnostic Interview-Revised; SCQ = Social Communication Questionnaire.

[1] Data were available for a subsample of n=235 infants.
[2] Data were available for a subsample of n=239 infants.
[3] Significant difference per class with $p<0.001$.
[4] Significant difference per class with $p<0.05$.

Clinical outcome was mixed in all identified classes. The distribution of outcome in each class is reported in Table 1. Although HR-ASD development was not specific to any class, a χ^2 test revealed a significant relationship between class membership and clinical outcome ($\chi^2(6)=13.39$; $p=0.037$). In particular, there were significantly more HR-ASD siblings in Class 1 compared to the other classes (odd ratio for HR-ASD in Class 1 compared to Class 3: OR=4.40 [CI: 1.90; 12.98], $p<0.001$). Although classes did not differ in ADOS CSS-Tot ($p=0.17$) nor in ADOS domain scores at 36 months (CSS-SA: $p=0.35$; CSS-RRB: $p=0.12$), differences were significant in ADI-Comm ($F(2,236)=9.56$, $p<0.001$), ADI-Soc ($F(2,236)=13.0$, $p<0.001$), ADI-RBI ($F(2,236)=5.32$, $p=0.005$) and SCQ-Tot scores at 36 months ($F(2,236)=11.19$, $p<0.001$). Post-hoc comparisons showed higher symptom severity for infants in Class 1 compared to the other classes ($p<0.001$ for all scores except for ADI-RBI showing $p<0.05$), while differences were not significant between Class 2 and Class 3 (Table 2). Thus, class membership significantly related to ASD symptoms and clinical outcome, with higher symptom severity and a higher risk for ASD in Class 1.

DISCUSSION

This study explored latent trajectories of adaptive functioning in infants at high and low familial risk for ASD. We observed variability in the development of adaptive functioning before age 2 and found three latent classes: one class with scores at or above age-appropriate norms at the first visit but decreasing trajectories afterwards (Class 1); one class with a relatively stable trajectory around age-appropriate norms (Class 2); and one class with increasing scores from below age-appropriate norms before age 2 to stable average scores afterwards (Class 3). Thus, high adaptive skills early in development were counterintuitively associated with poorer adaptive functioning in toddlerhood, while an initially delayed development appeared to be recovered by age 2. From age 2 onwards the identified classes mainly showed one relatively stable trajectory around average scores and a below-average decreasing trajectory, consistent with previous findings on older children with ASD[191-193]. Of note, classes significantly differed in ASD symptom severity and clinical outcome. Infants in Class 1 had significantly higher symptom severity at 36 months and there were more siblings who later met criteria for ASD than expected by a randomly distributed ASD outcome. Another important finding was a partial split between adaptive behaviour and cognitive development of the identified classes. In fact, Class 3 showed a stable trajectory around average cognitive level while it was identified by improving adaptive behaviour before age 2 and a stable trajectory around average scores afterwards.

Classes did not clearly map to clinical outcome groups. In our sample, 68% of infants developing ASD showed rather stable trajectories of adaptive skills around age-appropriate norms (Class 2), 23% of them showed decreasing skills (Class 1), and only 9% of them showed recovering adaptive skills by age 2 (Class 3). Similarly, the HR-Atypical outcome group was spread over the 3 classes, although the majority was in Class 2. This provides further support to the high heterogeneity of ASD in its phenotypic manifestations[30, 94, 209]. Yet, it is in contrast with our previous work on group comparisons, showing significantly lower adaptive skills in HR-ASD siblings compared to HR-Typical or LR controls and overall stable-low or decreasing trajectories of adaptive behaviour in infancy[55, 59]. However, our approach here focused on the identification of latent classes of trajectories independently from diagnostic outcome. Such approach allowed us to explore individual differences in early development and to identify different profiles within the ASD group.

Different trajectories of adaptive behaviour did not correspond to significant differences in ADOS scores, but there were significant differences in ADI-R and SCQ scores. This extends to infancy previous findings on discrepancies between adaptive behaviour and ADOS scores in pre-schoolers with ASD[191]. However, not all ASD symptoms are captured by the ADOS, and some children with “atypical” outcomes had by definition high ADOS scores (but did not meet diagnostic criteria for ASD). Rather, the relationship between adaptive behaviour and symptom severity can be captured by investigating all the instruments employed to assess symptom severity at 36 months (i.e. ADOS, ADI-R and SCQ). The split between ADI-R/SCQ and ADOS might be due to the parent-reported nature of ADI-R and SCQ scores, the same as VABS scores used to identify classes of infants, while ADOS scores are based on clinical observations.

It is remarkable to observe initially average or above-average adaptive functioning in some infants with later ASD outcome (Class 1). Nevertheless, they also showed a decreasing trajectory of scores compared to age-appropriate norms over time, which is in line with previous findings on children with ASD[191, 192]. Furthermore, the timing of the decline observed here is consistent with the emergence of overt behavioural signs of ASD as generally observed in previous studies around the second year of life[31, 32, 135]. Thus, an initial high level of adaptive functioning seems not to prevent ASD development; rather, when followed by a decline over time, it seems to be associated with a higher likelihood of ASD in toddlerhood. However, given that Class 1 represents only 7% of infants in our sample, we must be cautious with interpretations. Our findings are somewhat similar to what is observed in neuroimaging studies. There, higher fractional anisotropy and volume in the development of fiber tracts[210, 211], accelerated expansion of cortical surface area and brain volume overgrowth[112] in the first year of life were linked to later ASD outcome. One may speculate that high levels of adaptive skills early in development is linked to early alterations of brain development. In particular, hyper-expansion of cortical surface area may compromise the development of proper neural connectivity[212] and have downstream effects on behaviour, leading to the emergence of symptoms characteristic of ASD in toddlerhood. A recent framework for neurodevelopmental disorders suggests, in fact, that good synaptic compensation to overcome initial impairment at an earlier developmental stage might have more severe consequences later in development[213]. The declining trajectory of adaptive skills that we observed likely reflects little gain of skills between the first and second year of life and a failure to keep up with development. Further research should integrate more biological data on brain structure and brain function, such as EEG or MRI, to investigate what could have triggered the plateauing of skills for infants in Class 1. This would also provide more insight into the biological mechanisms underlying the identified developmental trajectories. Furthermore, while Class 2 and 3 were balanced in gender, Class 1 was mostly composed by males. This is consistent with the gender bias in prevalence of ASD[214] and with previous findings in children with ASD showing higher daily living scores in females than males[215]. Overall, this suggests that sex might moderate how

clinical symptoms are expressed in adaptive behaviour.

In terms of developmental level, the three classes of adaptive functioning correspond to two main classes of cognitive development. Infants falling behind age-appropriate norms in adaptive behaviour (Class 1) also showed decreasing trajectories in cognitive development, with significantly lower scores compared to their peers by age 2. On the other hand, infants with stable or improving adaptive behaviour (Class 2 and 3) did not differ in terms of developmental level, showing rather stable trajectories around age-appropriate norms. Discrepancy between adaptive and cognitive skills has been found before[59, 216], showing lower and more divergent adaptive skills compared to cognitive level. Our findings suggest that the main differences between decreasing and stable/increasing trajectories in adaptive behaviour might be driven by differences in cognitive level. Less cognitively able individuals appear to fall behind age-appropriate norms in adaptive behaviour while more cognitively able individuals show stable or increasing trajectories of adaptive behaviour around age-appropriate norms. Cognitive impairment has been shown to have a negative effect on everyday functioning and the development of adaptive behaviour, even beyond the effect of ASD symptoms[200]. However, the profile of good cognitive abilities with poor adaptive skills, which has been reported before in older children with ASD[199, 201], has not emerged from our study. Such profile might emerge later in life, as we have found in a follow-up of part of the current sample in mid-childhood[14]. This highlights the complexity of the development of adaptive behaviour and its relationship with cognitive development and ASD symptomatology, suggesting the necessity to further investigate adaptive behaviour in infancy.

Although findings are consistent, our study differs in three ways from the recent study of Sacrey et al.[198] on latent trajectories of adaptive behaviour in infancy. First, we identified classes across domains of adaptive behaviour instead of ABC score alone, providing more detailed profiles. Second, we used growth-mixture modelling[125] instead of latent class growth curve modelling[123], allowing variation within class to capture heterogeneity across individuals. Third, we examined younger ages by including data at 8 months. This earlier observation added particular value to our findings. In fact, children who later met criteria for ASD might not simply follow a trajectory of progressive impairment in adaptive skills, but some of them might present even stronger skills in the first year of life compared to other subgroups of infants. Our study has strengths in the relatively large sample of infant siblings (n=240) and the analysis of multiple instruments (MSEL, VABS, ADOS, ADI-R and SCQ), but it also has limitations. First, although trajectories were separated at the group level, there was still substantial overlap between classes in individual variation. Second, the three-step approach we used to examine the association of class membership with external variables might underestimate such relationship[217]. This is particularly true when classification errors in the assignment of individual class membership are high, and it might explain the discrepancy between VABS, MSEL and ADOS found in our study. However, high average class posterior probability was a selection criterion for the best model, reducing the impact of classification errors. Third, adaptive behaviour was assessed on the basis of parent-reported measures; however, class comparisons in terms of cognitive level and symptom severity, assessed by observational measures, enhance our confidence in the identified trajectories. Fourth, follow-up studies should investigate trajectories on a higher number of time-points to improve the estimate of the shape of the trajectory curve and test whether they change at later age. Finally, experimenters' awareness of risk-group status might lead to ascertainment bias due to more intense surveillance for ASD outcome among high-risk siblings compared to low-risk controls.

CONCLUSIONS

High-risk siblings and low-risk controls could be separated into three latent classes representing declining, improving and stable trajectories of adaptive behaviour between 8 and 36 months. We observed a dissociation between adaptive behaviour and cognitive development, with the improving class in adaptive behaviour showing stable trajectories of cognitive development around average scores. Furthermore, classes significantly differed in ASD symptoms and clinical outcome at 36 months. High levels of adaptive functioning in the first year of life followed by a failure to keep up with age appropriate norms was linked to higher symptom severity across the social, communication and repetitive behaviours domains. Furthermore, it was linked to increased likelihood of meeting diagnostic criteria for ASD in toddlerhood. Our findings provide better insight into the variety of paths leading to different functional outcomes within ASD. The identified subgroups indicate homogeneous classes of infants in terms of progression of adaptive functioning over time. These subgroups might be more relevant target groups for intervention aimed at improving later functioning.

5

SUPPLEMENTAL MATERIAL

TABLE S1. Descriptive statistics by risk group.

	Low-risk controls		High-risk siblings	
	mean	sd	mean	sd
Age				
8m	7.8	1.4	8.2	1.2
14m	14.3	1.3	14.5	1.3
24m	24.2	0.9	25.4	2.0
36m	38.3	2.6	39.0	3.0
VABS				
Comm 8m	101.7	13.6	93.6	16.2
Comm 14m	101.5	9.8	94.5	14.0
Comm 24m	109.4	11.8	101.1	12.7
Comm 36m	108.5	10.5	97.9	14.3
DL 8m	102.5	13.6	99.0	13.6
DL 14m	99.1	11.1	93.5	13.6
DL 24m	109.6	11.3	103.8	13.2
DL 36m	108.6	7.9	100.8	14.0
Mot 8m	96.3	13.9	86.8	16.5
Mot 14m	104.2	11.1	98.6	13.2
Mot 24m	103.6	9.0	98.5	11.3
Mot 36m	99.3	10.7	91.4	12.3
Soc 8m	102.8	12.6	98.8	12.7
Soc 14m	100.6	10.7	96.4	11.9
Soc 24m	106.8	10.6	98.2	11.0
Soc 36m	105.2	8.0	94.6	13.3
MSEL				
ELC 8m	107.4	12.6	102.7	15.8
ELC 14m	105.4	15.1	95.5	15.5
ELC 24m	115.6	14.2	99.8	20.2
ELC 36m	117.4	15.6	103.4	24.1
ADOS at 36m[1]				
CSS-Tot	2.64	2.0	3.09	2.56
CSS-SA	3.31	2.26	3.45	2.62
CSS-RRB	3.67	2.46	4.57	2.63
ADI-R at 36m[2]				
ADI-Comm	1.44	1.49	4.04	4.69
ADI-Soc	1.48	1.42	3.78	4.93
ADI-RBI	0.41	0.64	1.48	2.26
SCQ at 36m[2]				
SCQ-Tot	2.88	2.35	6.40	6.90

TABLE S1 CONTINUED.

	Low-risk controls		High-risk siblings	
	mean	sd	mean	sd
Gender				
Female	40 (54%)		82 (49%)	
Male	34 (46%)		84 (51%)	

Note. This table shows descriptive statistics by risk group for the entire sample.

Abbreviations. VABS = Vineland Adaptive Behavior Scales; Comm = communication score; DL = daily living score; Mot = motor score; Soc = socialization score; MSEL = Mullen Scales of Early Learning; ELC = early learning composite score; ADOS = Autism Diagnostic Observation Schedule; CSS = calibrated severity score; ADI-R = Autism Diagnostic Interview-Revised; ADI-Comm = Communication domain score (ADI-R); ADI-Soc = Social domain score (ADI-R); ADI-RBI = Restricted Behaviors and Interests domain score (ADI-R); SCQ = Social Communication Questionnaire; SCQ-Tot = Total score (SCQ);

[1] Data were available for a subsample of n=235 infants.

[2] Data were available for a subsample of n=239 infants.

TABLE S2: Model fitting.

	c	BIC	AIC	loglik	Pr	n1	n2	n3	n4	n5	n6
Common/diagonal variance matrix											
	1l	28665.76	28610.07	-14289.0	-	240	-	-	-	-	-
	1q	28593.31	28530.66	-14247.3	-	240	-	-	-	-	-
	2l	28610.63	28534.06	-14245.0	0.87	73	167	-	-	-	-
	2q	28528.49	28438	-14193.0	0.85	100	140	-	-	-	-
	3l	28608.76	28511.3	-14227.6	0.91	99	6	135	-	-	-
	3q	28522.89	28404.55	-14168.3	0.90	88	149	3	-	-	-
	4l	28620.78	28502.44	-14217.2	0.88	7	108	118	7	-	-
	4q	28522.98	28376.79	-14146.4	0.92	8	172	3	57	-	-
	5l	28634.85	28495.62	-14207.8	0.86	17	8	104	105	6	-
	5q	28546.99	28372.96	-14136.5	0.86	8	27	3	137	65	-
	6l	28646.49	28486.38	-14197.2	0.90	141	3	10	48	6	32
	6q	28582.72	28380.84	-14132.4	0.89	1	83	38	3	107	8
Common/not diagonal variance matrix											
	1l	28583.86	28517.73	-14239.9	-	240	-	-	-	-	-
	1q	28474.11	28390.58	-14171.3	-	240	-	-	-	-	-
	2l	28571.96	28484.94	-14217.5	0.95	15	225	-	-	-	-
	2q	28470.90	28359.52	-14147.8	0.95	227	13	-	-	-	-
	3l	28587.99	28480.09	-14209.0	0.88	39	189	12	-	-	-
	3q	28485.11	28345.88	-14132.9	0.97	13	224	3	-	-	-
	4l	28608.42	28479.64	-14202.8	0.90	12	185	3	40	-	-

TABLE S2 CONTINUED.

	c	BIC	AIC	loglik	Pr	n1	n2	n3	n4	n5	n6
Common/not diagonal variance matrix											
	4q	28505.28	28338.21	-14121.1	0.93	189	37	3	11	-	-
	5l	28631.39	28481.72	-14197.9	0.90	74	3	32	119	12	-
	5q	28544.80	28349.88	-14118.9	0.93	185	1	15	36	3	-
	6l	28649.22	28478.67	-14190.3	0.90	42	64	6	111	5	12
	6q	28569.67	28346.91	-14109.5	0.87	66	4	13	11	31	115
Class-specific/diagonal variance matrix											
	1l	28665.76	28610.07	-14289.0	-	240	-	-	-	-	-
	1q	28593.31	28530.66	-14247.3	-	240	-	-	-	-	-
	2l	28601.35	28521.3	-14237.6	0.88	163	77	-	-	-	-
	2q	28522.85	28428.87	-14187.4	0.87	135	105	-	-	-	-
	3l	28605.06	28500.64	-14220.3	0.89	157	10	73	-	-	-
	3q	28523.12	28397.82	-14162.9	0.87	20	177	43	-	-	-
	4l	28612.39	28483.61	-14204.8	0.90	135	11	88	6	-	-
	4q	28524.58	28367.95	-14139.0	0.85	134	37	4	65	-	-
	5l	28638.78	28485.63	-14198.8	0.84	88	105	30	6	11	-
	5q	28556.29	28368.34	-14130.2	0.83	117	6	30	79	8	-
	6l	28666.19	28488.68	-14193.3	0.86	45	129	42	11	6	7
	6q	28591.30	28372.01	-14123.0	0.85	123	34	19	4	59	1
Class-specific/not diagonal variance matrix											
	1l	28583.86	28517.73	-14239.9	-	240	-	-	-	-	-
	1q	28474.11	28390.58	-14171.3	-	240	-	-	-	-	-
	2l	28578.27	28487.77	-14217.9	0.83	177	63	-	-	-	-
	2q	28475.58	28360.72	-14147.4	0.85	188	52	-	-	-	-
	3l	28599.08	28484.22	-14209.1	0.85	119	85	36	-	-	-
	3q	28484.42	28338.23	-14127.1	0.90	171	65	4	-	-	-
	4l	28616.55	28477.32	-14198.7	0.90	41	12	182	5	-	-
	4q	28521.22	28343.71	-14120.9	0.93	68	2	168	2	-	-
	5l	28646.30	28482.71	-14194.4	0.81	159	38	27	4	12	-
	5q	28557.60	28348.76	-14114.4	0.89	169	24	5	41	1	-
	6l	28681.00	28493.05	-14192.5	0.89	36	161	23	2	6	12
	6q	28602.14	28361.98	-14112.0	0.85	5	24	115	44	41	11

Note. This table shows the metrics of model fitting for the different models tested based on the polynomial degree of the growth curve, the variance/covariance matrix and the number of classes. Separate sections illustrate results for models differing in variance/covariance matrices across classes (common vs class-specific, and diagonal vs non-diagonal matrices). Each section shows results for linear and quadratic growth with 1 to 6 classes. The 3-class quadratic model with class-specific and diagonal variance/covariance matrix for random effects was the one selected as best model for further analysis.

Abbreviations. c = number of classes; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; loglik = log-likelihood; Pr = average class posterior probability; n1-n6 = number of infants in each class.

6

Uncovering Neuro-developmental Paths to Autism Spectrum Disorder through an Integrated Analysis of Developmental Measures and Neural Sensitivity to Faces

Submitted.

Bussu, G.,
Llera, A.,
Jones, E. J. H.,
Tye, C., Charman,
T., Johnson,
M. H., Beckmann,
C.F., Buitelaar,
J. K., & The BASIS Team

ABSTRACT

BACKGROUND

Autism Spectrum Disorder (ASD) is characterized by high heterogeneity in etiology and manifestation. The neurobiological processes underlying ASD development are reflected in multiple features, from behavior and cognition to different indices of brain functioning. An integrated analysis of these features may optimize identification of these processes and maximize inter-individual variation.

METHODS

We examined neural sensitivity to eye-gaze at 8 months, then cognitive and adaptive functioning, and early ASD symptoms between 8 and 36 months in infants at familial high-risk for ASD (HR, n=161) and low-risk controls (n=71). At 36 months, HR siblings were clinically classified as having typical development, atypical development, or ASD. We used linked independent component analysis to extract patterns of variation across domains and development, and selected the patterns significantly associated with clinical classification.

RESULTS

We identified two independent, longitudinal processes and one early stage process having coherent effects across multiple domains of development. The early process at 8 months was associated with non-ASD outcome and characterized by high functioning and low levels of ASD symptoms linked to higher attention to gaze shifts. One longitudinal process was associated with non-ASD outcome and characterized by increasing functioning and low levels of ASD symptoms, while the other indicated a stagnation in cognitive functioning at 24 months and was associated with ASD outcome.

CONCLUSIONS

Results highlight the complexity of emerging ASD, which goes beyond the limits of clinical categories. The processes identified deepen our understanding of the underlying neurodevelopmental mechanisms associated with emerging ASD. Future work could link them to independent suites of genes, providing insight into the specific genetic risks for that phenotype.

INTRODUCTION

Autism Spectrum Disorders (ASD) are behaviourally defined by difficulties in social-communication, and the presence of restricted and repetitive patterns of symptoms and interests and sensory anomalies (APA, 2013). The intrinsic heterogeneity of ASD is evident at different levels of analysis and points to multiple underlying biological mechanisms leading to the disorder[30, 98]. This multiplicity will likely not be captured by focusing on a single functional domain as coherent biological processes may ultimately express themselves across multiple domains of functioning. Thus, integration of information from different domains might be essential to uncover these underlying processes. Furthermore, the investigation of longitudinal measurements can help understand how these processes express themselves across domains over time, as heterogeneity is also observed within the same individual throughout development[96]. This study aimed to uncover underlying processes early in development linked to later emergence of ASD. To do that, we looked for coherent patterns of variation across multiple developmental domains over time through an integrated analysis, in contrast to previous studies that have reported on categorical analyses that were only post-hoc associated across domains.

Prospective longitudinal studies of infants at high familial risk for ASD (HR), based on having an older sibling with ASD, allow us to study the early manifestations of the disorder by investigating differences between those infants who develop ASD and those who don't[29]. Infants developing ASD demonstrate emerging atypicalities in social-communicative behaviour from the first year of life, with a declining interest in human faces[37, 80, 81] by 6 to 12 months of age. Event-related potentials (ERPs) provide a useful tool to examine the neural correlates of face recognition in infancy[71]. The ERP waveform in response to faces in infancy typically presents the characteristic P1, N290, and P400 components, known to be modulated by direction of eye-gaze as early as 4 months of age[77]. HR siblings developing ASD have shown a more rapid P400 peak response to faces versus objects at 6 months compared to their typically developing peers[86], and reduced differentiation in P400 amplitude responding to faces that shift gaze towards versus away from the viewer compared to non-ASD siblings between 6 and 9 months[87]. Similarly, atypical neural responses to social stimuli in high-risk siblings were associated to difficulties in socialisation in toddlerhood[88]. Previous studies have also reported behavioural signs of ASD in the first two years of life, including reduced fixation to the eye region between 2 and 6 months[218], reduced gaze fixation to people at 6 months[38], impairments in verbal[56] and motor skills at 7 months[53], and lower developmental level by 14 months, with significantly lower scores on all scales of the Mullen Scales of Early Learning (MSEL) except for Visual Reception[49]. In particular, previous longitudinal studies on high-risk siblings for ASD have

shown atypicalities across measures of developmental level (MSEL), adaptive functioning (Vineland Adaptive Behavior scales, VABS) and early ASD symptoms (Autism Observation Scales for Infants, AOSI) emerging in the sensorimotor domain at 6 months and in the social-communication domain after 12 months[35, 55]. Overall, there is a general consensus in the field that the defining behavioural features of ASD are not present in the first year of life but begin to emerge around 12 months and consolidate between 18 and 36 months[212, 219]. However, this pre-symptomatic period is characterised by sensorimotor[35, 40, 55] and visual attention[38, 220-222] atypicalities and by alterations in brain structure[112, 223, 224] and function[86, 87, 111] in infants with later ASD outcome.

Although valuable to identify potential early risk markers for ASD, the traditional supervised approach is based on predefined clinical labels to partition the sample, treating prospective data as a case-control design by assuming that the clinical category has a meaning at the start. However, such an assumption might not be valid and it does not take into account the heterogeneity of the resulting clinical groups, which often overlap across symptoms[121]. Unsupervised data-driven methods allow the identification of intrinsic patterns in data without any a priori knowledge on how the different measures relate to each other or to clinical outcome. Here, we employed unsupervised learning methods to separate underlying neurodevelopmental processes associated with clinical outcome based on the extraction of intrinsic patterns in multivariate unlabelled data. As opposed to the more traditional retrospective investigation of early differences between categorical outcomes, our approach to prospective analysis allows to understand the different emerging patterns of development and how they lead to specific outcomes by looking at structure in the data. The identified patterns might then be the key to improve our understanding of individual heterogeneity and allow stratification into more homogeneous and predictable subgroups, that might be a better target for early intervention.

Linked independent component analysis (ICA) can be used to simultaneously model and discover common features across multiple modalities[126, 127]. Although mainly used in neuroimaging[130-132], this method can be directly applied to any type of multimodal data acquired for a fixed group of participants. Thus, it can be applied to model longitudinal multimodal data collected from large cohorts of infant siblings and identify underlying biological processes with expression in different domains across development. In this study, we used linked ICA to uncover neurodevelopmental processes acting early in development by simultaneous factorization of developmental measures and electrophysiological measures of neural sensitivity to social and non-social stimuli at 8 months. The same approach was used to uncover underlying processes acting across development by simultaneous factorization of longitudinal measures of developmental level, adaptive functioning, and emerging ASD symptoms between 8 and 36 months. Then, we tested in a post-hoc analysis the association of the identified processes to clinical outcome at 36 months. This provided novel insights into the neurodevelopmental processes acting together from early age and leading to different clinical outcomes depending on their presence at an individual level.

6

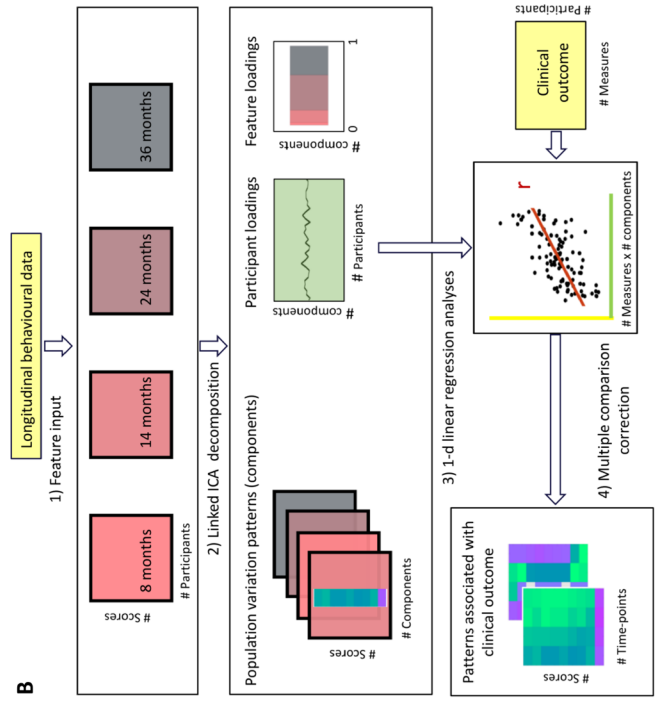
METHODS AND MATERIALS

We performed two separate analyses (Figure 1): a multimodal analysis to identify early neurodevelopmental patterns, and a longitudinal analysis to identify processes acting across development.

PARTICIPANTS

Data were collected from infants recruited in one of two phases of the British Autism Study of Infant Siblings (BASIS, http://www.basisnetwork.org)[87, 225], involving low-risk controls (LR) and high-risk siblings (HR). All procedures were in agreement with ethical approval granted by the London Central NREC (approval codes 06/MRE02/73, 08/H0718/76), and one or both parents gave informed consent to participate in the study. Experimenters were aware of infants' risk status, but assessments were blind to clinical outcome. At the time of enrolment, none of the infants had been diagnosed with any developmental condition. The longitudinal sample included 232 infants (71 LR controls and 161 HR siblings) followed on four visits at 8.1±1.2 months (mean ± standard deviation; hereafter 8 months), 14.5±1.3 months (hereafter 14 months), 25.4±3.1 months (hereafter 24 months) and 38.4±2.3 (hereafter 36 months). To handle missing data, we performed imputation through expectation maximization on SPSS (http://www.ibm.com/analytics/us/en/technology/spss, see *Supplemental Material* for details). The multimodal analysis was run in a subsample of 201 infants (61 LR controls and 140 HR siblings) selected because of having neural data available at 8 months (8.14±1.22 months). Both samples were balanced in gender (see Table 1).

FIGURE 1: Analysis flowchart ***[Figure on the next page]*****.** This figure illustrates the different steps of analysis for the extraction of underlying processes associated with clinical outcome at 36 months. Panel A illustrates the multimodal analysis, integrating clinical and ERP data both collected at 8 months, while panel B illustrates the longitudinal analysis, integrating behavioral data from standardized clinical instruments collected between 8 and 36 months.



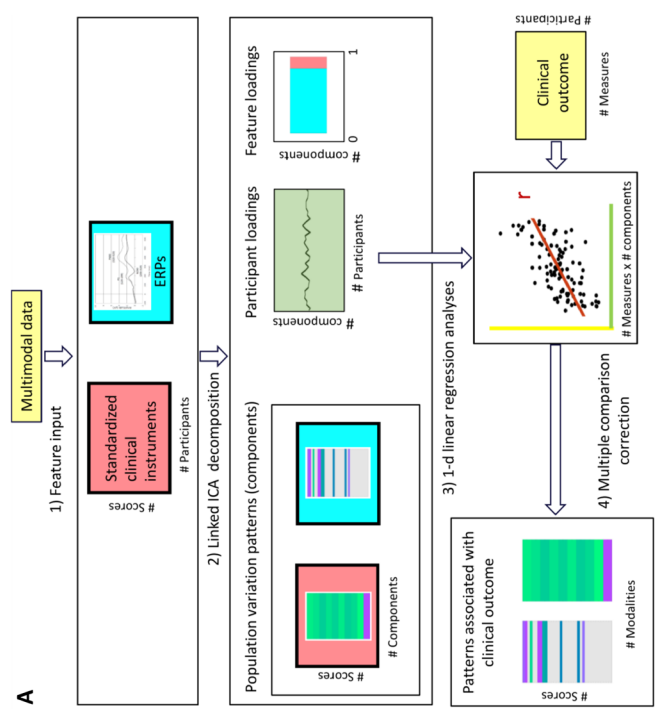


FIGURE 1: Analysis flowchart

MEASURES

Developmental skills

Cognitive development was measured at each visit by the Mullen Scales of Early Learning (MSEL[203]), a standardized developmental measure assessing cognitive functioning in 5 scales: gross motor (GM), visual reception (VR), fine motor (FM), receptive (RL) and expressive language skills (EL). T-scores (mean=50; standard deviation, SD=10) from the 5 scales at 8 months were included as input features in the multimodal analysis. Gross motor scores were excluded from the longitudinal analysis as not available at 36 months, leading to 4 input features from the MSEL.

Adaptive functioning

Adaptive behavior was measured by the Vineland Adaptive Behavior Scales (VABS-II[196]), a semi-structured parent-report questionnaire (at 8 and 14 months) or parent interview (at 24 and 36 months) assessing personal and social functioning in 4 different domains: Communication (Comm), Daily Living Skills (DL), Socialization (Soc) and Motor Abilities (Mot). Standard scores (mean=100; SD=15) from the 4 domains were included as input features in all analyses.

Early ASD symptoms

A 19-item version of the Autism Observation Scale for Infants (AOSI), a semi-structured observational assessment[138], was administered at 8 and 14 months to detect putative behavioural signs of ASD. The AOSI total score at 8 months was used as input feature in the multimodal analysis. To assess ASD symptomatology, the Autism Diagnostic Interview Revised (ADI-R[7]) was administered at 36 months and the Autism Diagnostic Observation Schedule (ADOS-2[204]) was administered at 24 and 36 months. Item scores from the ADOS-G were used to calculate ADOS-2 total scores. All LR siblings were administered module 2; 138 HR were administered module 2 and 23 HR were administered module 1.
Total scores from the AOSI at 8 and 14 months, and from the ADOS at 24 and 36 months were used as input features for the longitudinal analysis.

Event-Related Potentials (ERPs)

The task was the same as in Elsabbagh et al[87]. It was designed to assess responses to: static face [Fc]; visual noise stimuli [Ns]; static faces with direct [FD] gaze; static faces with averted gaze [FA]; gaze shifts toward the infant [SD]; gaze shifts away from the infant [SA]. Components P100, N290, and P400 averaged across occipito-temporal channels were quantified by amplitude and latency for a total of 36 ERP variables measured at 8 months and used as input features for the multimodal analysis (see *Supplemental Material* for details).

Clinical outcome evaluation

The LR group was based on having an older full sibling with typical development. LR infants received no formal clinical diagnoses, but none of them had a community clinical ASD diagnosis at 36 months. In particular, no ADI-R was administered to LR in Phase 1, who did not receive an outcome evaluation. In Phase 2, LR infants were administered the ADOS and ADI-R and received an outcome evaluation at 36 months, but none of them raised any concern for ASD or atypical development. HR siblings received a clinical outcome evaluation at 36 months and were subsequently grouped into siblings with ASD (HR-ASD); with atypical (non-ASD) development (HR-Atypical); and with typical development (HR-Typical).

Expert clinical researchers reviewed all available information at 24 months and 36 months and assigned clinical consensus best estimate diagnosis of ASD according to ICD-10[9] or DSM-5 criteria[8], depending on the phase of the study. The best estimate diagnoses for the two phases were reviewed for differences in categorization and considered to be similar. Criteria for 'atypical' categorization included: ADOS and/or ADI-R above ASD threshold, and/or MSEL more than 1.5 standard deviations below average on visual reception and/or receptive language and/or expressive language and/or early learning composite (see *Supplemental Material* for details).

Statistical Analysis

Linked ICA is a Bayesian extension of independent component analysis (ICA) for unsupervised learning of statistically independent modes of variation in data[127, 128], allowing for the simultaneous analysis of multimodal data collected on the same participants[126]. The identified components indicate processes considered independent based on how they affect different measures (i.e. across behavioural or neural data), but linked across modalities (i.e. behavioural versus brain data, Figure 1A) or time-points (Figure 1B). Each component explains variation within the individual participant (Figure 1A.2, 1B.2) and is represented by: (1) a vector of individual loadings, namely scalar values indicating how much that component explains developmental variation for the individual participant; (2) component weightings in different modalities (i.e. time-points in the longitudinal analysis; behavioural and neural data in the multimodal analysis); (3) a score map, indicating the relative value of scores compared to the estimated noise in individual variability. For the implementation, we used the code available on the FSL homepage (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA). The number of independent components was estimated such that more than 90% of variance was explained.

In the multimodal analysis, we integrated developmental data (n=10 total features from

MSEL, VABS and AOSI) and ERP data at 8 months (n=36 total features; Figure 1A.1). The number of components was estimated to be 10. In the longitudinal analysis, we integrated developmental data (n=9 total features from MSEL, VABS and AOSI/ADOS) between 8 and 36 months. Different time-points were considered as different input modalities (Figure 1B.1) but were not considered as ordinal. The number of components was estimated to be 9.

To evaluate whether the extracted components were related to clinical outcome, we computed the association of the individual loadings to clinical outcome at 36 months through linear regression (Figure 1A.3, 1B.3). Clinical outcome was included as dependent variable, while individual component loadings and gender were included as dependent variables. We used Bonferroni correction for multiple comparisons by correcting for the number of components (Figure 1A.4, 1B.4). When the main effect of clinical outcome was significant, post-hoc differences between clinical outcome groups were tested through t-tests in robust ranges, and group differences were considered significant for $p<0.05/6=0.008$ (tests=6). Differences in competence at different time-points, computed as average of MSEL and VABS scores, were also tested via t-tests.

RESULTS

Data

Demographics and clinical characteristics of the two samples are shown in Table 1, Table 2 and Table 3. Clinical outcome groups did not differ in age at any visit, while gender was significantly different per clinical outcome ($\chi^2(3)=11.55$, $p=0.009$ in the multimodal analysis; $\chi^2(3)=9.66$, $p=0.022$ in the longitudinal analysis), with more males among HR-ASD.

TABLE 1. Demographics

		Overall	HR-ASD	HR-Atypical	HR-Typical	LR
Longitudinal analysis						
		n	n	n	n	n
		232	32	43	86	71
Gender*						
	Male	118	24	23	38	33
	Female	114	8	20	48	38
Age		mean (SD)	mean (SD)	mean (SD)	mean (SD)	mean (SD)
	8 m	8.13 (1.22)	8.03 (1.12)	8.33 (1.06)	8.24 (1.21)	7.92 (1.35)
	14 m	14.48 (1.27)	14.50 (1.32)	14.56 (1.20)	14.58 (1.29)	14.31 (1.26)
	24 m	25.39 (3.06)	24.84 (1.63)	26.40 (4.25)	25.72 (2.31)	24.63 (3.30)
	36 m	38.39 (2.32)	38.06 (1.90)	38.19 (2.05)	38.62 (2.29)	38.39 (2.69)
Multimodal analysis						
		n	n	n	n	n
		201	30	36	74	61
Gender**						
	Male	99	23	18	30	28
	Female	102	7	18	44	33
Age		mean (SD)	mean (SD)	mean (SD)	mean (SD)	mean (SD)
	8 m	8.14 (1.22)	8.03 (1.05)	8.31 (1.09)	8.27 (1.16)	7.92 (1.41)

Note: This table shows gender (count, *n*) and age by clinical outcome group.

Abbreviations: ASD= autism spectrum disorder; HR = high-risk siblings; LR = low-risk controls.

* Significant difference of gender per clinical outcome: $\chi^2(3) = 9.66$, $p = 0.022$.

** Significant difference of gender per clinical outcome: $\chi^2(3) = 11.55$, $p = 0.009$.

TABLE 2. Clinical characteristics of the longitudinal sample

	Overall (n=232)				High-Risk (n = 161)												Low-Risk (n = 71)			
					ASD (n = 32)				Atypical (n = 43)				Typical (n = 86)							
	8 m	14 m	24 m	36 m	8 m	14 m	24 m	36 m	8 m	14 m	24 m	36 m	8 m	14 m	24 m	36 m	8 m	14 m	24 m	36 m
MSEL																				
GM	47.24 (10.68)	49.57 (14.79)	51.54 (11.50)	-	43.84 (11.37)	45.59 (14.46)	46.95 (14.02)	-	45.07 (12.57)	47.60 (13.01)	49.46 (10.60)	-	47.31 (10.48)	50.87 (14.97)	50.91 (11.39)	-	50.00 (8.68)	50.97 (15.54)	55.65 (9.72)	-
FM	55.10 (12.41)	55.69 (9.95)	48.77 (10.67)	51.40 (16.39)	48.53 (12.91)	50.50 (11.63)	44.64 (11.55)	39.84 (16.09)	52.00 (12.78)	52.63 (11.07)	45.17 (13.30)	43.30 (16.01)	56.77 (12.71)	56.50 (8.43)	48.49 (8.95)	54.34 (14.80)	57.9 (10.17)	58.89 (8.83)	53.15 (8.83)	57.96 (14.02)
VR	53.81 (11.07)	49.99 (9.98)	53.62 (12.85)	56.83 (13.82)	51.59 (10.46)	45.09 (9.31)	47.30 (13.08)	49.29 (17.54)	50.49 (11.16)	48.53 (9.56)	46.92 (13.62)	49.47 (15.04)	54.35 (11.75)	48.95 (9.96)	55.18 (10.59)	60.51 (10.88)	56.17 (9.95)	54.35 (9.05)	58.65 (12.14)	60.21 (11.29)
RL	47.56 (10.17)	42.64 (11.90)	52.17 (12.99)	52.53 (12.85)	43.35 (12.50)	36.19 (9.05)	41.71 (15.47)	43.47 (17.71)	46.05 (8.70)	40.19 (10.73)	46.92 (13.17)	43.37 (13.04)	49.90 (10.08)	43.68 (12.31)	53.26 (10.74)	55.80 (8.37)	47.55 (9.33)	45.77 (12.01)	58.75 (9.71)	58.21 (9.15)
EL	51.05 (10.21)	47.77 (10.66)	51.60 (12.93)	53.95 (12.86)	50.08 (11.89)	42.13 (11.44)	46.23 (15.30)	43.28 (16.14)	51.26 (11.00)	44.98 (10.90)	47.72 (13.20)	45.20 (12.31)	50.52 (10.14)	49.64 (9.93)	50.70 (11.49)	57.76 (9.33)	52.01 (9.07)	49.75 (9.93)	57.46 (11.18)	59.39 (9.38)
VABS																				
Comm	95.73 (15.88)	96.59 (13.30)	103.31 (12.73)	101.12 (14.28)	90.19 (15.33)	86.04 (14.28)	94.47 (15.06)	88.96 (18.19)	89.53 (17.12)	93.17 (15.85)	98.41 (10.78)	93.18 (14.03)	96.35 (15.71)	98.35 (11.63)	104.31 (10.66)	103.26 (10.24)	101.23 (13.58)	101.29 (9.68)	109.06 (11.88)	108.82 (10.42)
DL	99.95 (13.49)	95.17 (12.99)	105.48 (12.48)	103.06 (13.03)	93.56 (15.28)	85.63 (13.39)	97.88 (13.95)	88.36 (18.03)	98.98 (11.75)	93.38 (13.38)	101.49 (13.50)	97.74 (12.64)	101.24 (13.14)	96.58 (12.62)	107.55 (11.09)	106.63 (8.84)	101.87 (13.39)	98.86 (10.80)	108.83 (10.80)	108.60 (7.91)
Mot	89.65 (16.21)	100.33 (12.84)	99.97 (10.66)	93.66 (12.23)	85.13 (16.71)	98.06 (14.27)	98.19 (12.49)	84.66 (13.25)	80.65 (15.66)	95.73 (15.10)	96.51 (11.83)	86.47 (10.48)	90.73 (15.88)	100.40 (11.59)	99.50 (10.09)	95.79 (10.35)	95.84 (13.80)	104.04 (11.15)	103.44 (8.77)	99.49 (10.54)
Soc	99.84 (12.72)	97.77 (11.66)	100.72 (11.46)	97.78 (12.89)	96.97 (15.66)	91.40 (11.78)	88.91 (11.68)	79.64 (12.66)	97.93 (11.52)	96.53 (12.66)	97.33 (10.65)	92.43 (11.75)	99.62 (11.85)	98.44 (11.06)	101.99 (8.21)	100.93 (8.68)	102.55 (12.71)	100.58 (10.71)	106.56 (10.74)	105.38 (8.00)
AOSI	8.34 (4.86)	5.10 (4.34)			10.66 (5.71)	7.59 (4.42)			9.47 (4.86)	7.05 (4.94)			8.51 (4.66)	4.48 (3.98)			6.39 (3.96)	3.56 (3.41)		
ADOS			5.04 (5.04)	5.75 (4.84)			11.25 (5.59)	10.63 (6.54)			6.67 (5.07)	8.95 (4.95)			4.13 (3.72)	3.06 (2.04)			2.35 (3.16)	4.89 (3.42)

Note: This table shows measures [*mean (standard deviation, SD)*] from standardized clinical instruments by clinical outcome group.

Abbreviations: ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; GM= gross motor abilities (MSEL); FM= fine motor abilities (MSEL); VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); VABS = Vineland Adaptive Behavior Scales; Comm = communication skills (VABS); DL = daily living skills (VABS); Soc = social skills (VABS); Mot = motor skills (VABS); AOSI= Autism Observation Scale for Infants; ADOS= Autism Diagnostic Observation Schedule; HR = high-risk siblings; LR = low-risk controls.

TABLE 3. Clinical characteristics of the sample included in the multimodal analysis

	MSEL					VABS			AOSI
	GM	FM	VR	RL	EL	Comm	DL	Mot	Soc	Total score
LR (n=61)	49.49 (8.42)	57.21 (10.11)	55.48 (9.73)	47.72 (8.77)	51.90 (9.21)	100.1 (13.24)	100.85 (12.87)	94.85 (13.49)	101.38 (12.16)	6.36 (3.91)
HR-Typ (n=74)	46.86 (10.52)	55.91 (12.77)	54.42 (11.95)	49.53 (10.14)	50.20 (10.22)	96.16 (15.75)	100.85 (13.68)	89.86 (16.24)	99.97 (12.51)	8.82 (4.75)
HR-Atyp (n=36)	43.44 (12.58)	51.67 (12.76)	51.42 (11.51)	45.14 (8.80)	51.69 (11.43)	91.83 (15.40)	100.08 (11.22)	79.67 (15.95)	98.69 (10.45)	9.53 (4.86)
HR-ASD (n=30)	42.82 (10.47)	47.80 (11.97)	50.46 (9.32)	42.80 (12.23)	49.82 (11.83)	89.77 (15.73)	93.17 (15.42)	84.17 (16.26)	96.87 (16.18)	10.90 (5.52)

Note: This table shows measures [*mean (standard deviation, SD)*] at 8 months from standardized clinical instruments by clinical outcome group.

Abbreviations: ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; GM= gross motor abilities (MSEL); FM= fine motor abilities (MSEL); VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); VABS = Vineland Adaptive Behavior Scales; Comm = communication skills (VABS); DL = daily living skills (VABS); Soc = social skills (VABS); Mot = motor skills (VABS); AOSI= Autism Observation Scale for Infants; HR = high-risk siblings; LR = low-risk controls.

MULTIMODAL PATTERNS OF DEVELOPMENTAL AND ERP DATA

Among the linked components across behavioural and brain data at 8 months, only one component was significantly associated to clinical outcome at 36 months (IC7: β=-0.29, $p<0.001$; Figure 2). This was a multimodal component (Figure 2C) showing a pattern in ERP variables (Figure 2A) characterized by longer P1 latency in response to gaze shifting away; higher P400 amplitude, lower P1 amplitudes and shorter N290 latency in responses to gaze shifts towards and away from the infant; and lower P1 amplitude in response to visual noise. The linked pattern in clinical measures at 8 months showed high levels of competence across all functional domains and low level of early ASD symptoms (Figure 2B). In particular, scores were higher in gross motor, visual reception, and receptive language MSEL scores, and communication and motor VABS scores. Individual loadings were negatively associated to clinical outcome (β=-0.29, Figure 2D), meaning that the identified process was present more strongly in LR and HR-Typical infants. However, outcome groups were not clearly separated, with a significant difference only between LR and HR-Atypical ($p=0.005$) or HR-ASD ($p<0.001$). The effect of gender covariate was not significant (β=-0.37, $p=0.007$ with $\alpha_{Bonferroni}$=0.005). Detailed statistics can be found in Table 4.

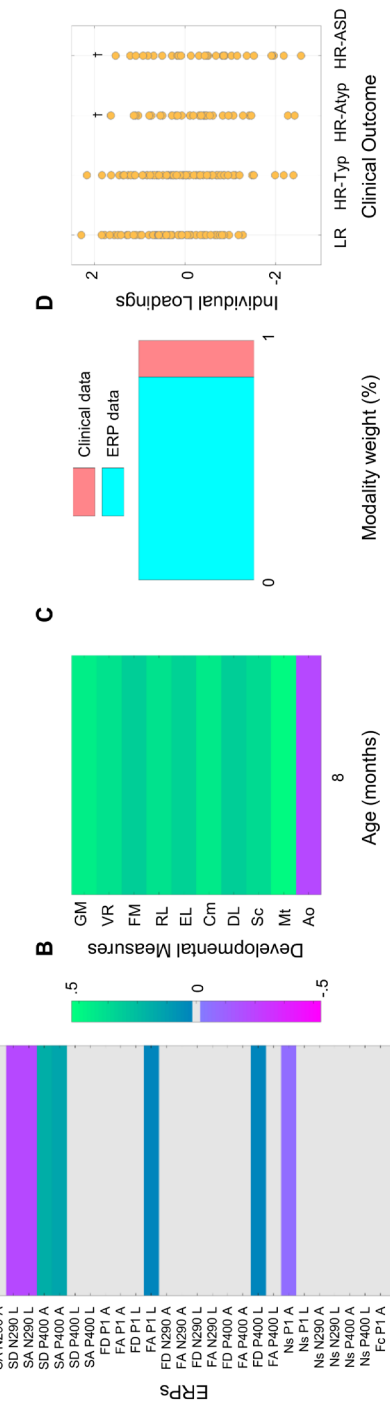


FIGURE 2. Independent component linked across modalities ***[Legend on the next page]***.

FIGURE 2. Independent component linked across modalities. This figure illustrates the independent component linked across ERP and clinical data, both collected at 8 months, significantly associated to clinical outcome at 36 months (IC7). Panels A and B respectively show the associated sources of variation for ERP and clinical scores. Panel C presents the contribution of each measure to the component and D shows individual participant loadings to the component grouped by clinical outcome at 36 months: significance (Bonferroni corrected p<.05) is indicated by ' for group differences from the LR group.

Abbreviations: SA = averted gaze shift; SD = direct gaze shift; FA = static averted gaze; FD = static direct gaze; Fc = face with static gaze (average between direct and averted); Ns = visual noise; A = amplitude; L = latency; GM = gross motor scores (MSEL); VR = visual reception scores (MSEL); FM = fine motor scores (MSEL); RL = receptive language scores (MSEL); EL = expressive language scores (MSEL); Cm = communication scores (VABS); DL = daily living scores (VABS); Sc = social scores (VABS); Mt = motor scores (VABS); Ao = AOSI total score; MSEL = Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scale; AOSI = Autism Observation Scales for Infants.

LONGITUDINAL PATTERNS OF DEVELOPMENTAL DATA

Among the linked components in longitudinal data from standardized clinical instruments, two components were significantly associated to clinical outcome at 36 months (IC1: β=-0.60, p<0.001; and IC3: β=0.22, p<0.001). The first component (IC1, Figure 3, top row) was characterized by increasing competence across domains of cognitive (MSEL scores) and adaptive functioning (VABS scores) between 8 and 36 months, reaching the peak in communication, daily living and social skills at 36 months, while the level of ASD symptoms (AOSI and ADOS scores) was low over time (Figure 3A). Development of competence increased significantly between 8 and 14 months (t(7)=-3.99, p=0.005), and between 14 and 24 months (t(7)=-8.25, p<0.001), while the increase between 24 and 36 months was not significant after Bonferroni correction (t(7)=-2.73, p=0.029) (Figure 3D). The identified process mostly explained variance from measures at 24 and 36 months (Figure 3B) and was negatively associated to clinical outcome (β=-0.60, Figure 3C), meaning that it was present more strongly in LR and HR-Typical infants. In fact, individual loadings on this component were higher in LR controls and HR-Typical than HR-Atypical and HR-ASD siblings (Figure 3C). Differences were significant (p<0.001) between all clinical outcome groups except for HR-Atypical and HR-ASD (see Table 4). The effect of gender covariate was not significant (β=-0.04, p=0.73 with $\alpha_{Bonferroni}$=0.006).

IC3 (Figure 3, bottom row) explained mostly variance on measures at 24 and 36 months (Figure 3F). It started with low levels of cognitive abilities at 8 months, followed by an increase in ASD symptoms severity, visual receptive abilities and motor abilities (MSEL fine motor and VABS motor scores) by 24 months, and by a further increase in severity of ASD symptoms and a plateau in cognitive and adaptive functioning at 36 months (Figure

3E). In particular, average competence across cognitive and adaptive functioning domains decreased significantly between 24 and 36 months ($t(7)=5.07$, Bonferroni corrected $p=0.004$; Figure 3H). As expected for low levels of functioning and high severity of ASD symptoms, this pattern of scores was positively associated to clinical outcome ($\beta=0.22$, Figure 3G), with increasing individual loadings from HR-Typical to HR-ASD. Thus, the observed process was present more strongly in HR-Atypical and HR-ASD siblings; however, low-risk controls had higher loadings than HR-Typical siblings. Differences among clinical outcome groups were all significant ($p<0.001$) except for LR vs. HR-Atypical and HR-Atypical vs. HR-ASD (see Table 4). Furthermore, there was a significant effect of gender covariate on clinical outcome ($\beta=-0.40$, $p=0.002$ with $\alpha_{Bonferroni}=0.006$). Thus, as opposed to the other identified components, individual scores on IC3 do not explain variance in clinical outcome over and above gender.

To test the link between the identified components from the two different analyses, we correlated the behavioural score map of IC7 (excluding gross motor scores) with behavioural score maps of IC1 and IC3 at 8 months (Figure 2B versus Figure 3A column 1, and Figure 2B versus Figure 3E column 1). Results showed a significant correlation between IC1 and IC7 ($r=0.92$, $p<0.001$), while correlation was not significant between IC3 and IC7 ($r=-0.13$, $p=0.73$).

FIGURE 3. Independent components linked across development *[Figure on the next page]*. This figure shows results for the independent components obtained from the analysis of longitudinal clinical data: IC1 (top row) and IC3 (bottom row). Panels A and E show the associated sources of variation for clinical scores at different time-points respectively for the two independent processes identified. Similarly, panels B and F present the contribution of each time-point to the components, while panels C and G show individual participant loadings to the components grouped by clinical outcome at 36 months: significance (Bonferroni corrected $p<.05$) is indicated by ' for group differences from the LR group and # for group differences from the HR-Typical group. Finally, panels D and H show the trajectories of average competence across all functional domains (VR, FM, RL, EL, Cm, DL, Sc, Mt) respectively for the two independent processes identified. The red line marks the median of scores as shown in panels A and E, and indicates: (D) a significant increase in average competence between 8 and 14 months (*, $p<.05$), reaching its peak at 24 months (***, $p<.001$); (H) a significant decrease in average competence between 24 and 36 months (**, $p<0.005$).

Abbreviations: VR = visual reception scores (MSEL); FM = fine motor scores (MSEL); RL = receptive language scores (MSEL); EL = expressive language scores (MSEL); Cm = communication scores (VABS); DL = daily living scores (VABS); Sc = social scores (VABS); Mt = motor scores (VABS); Ao = ASD symptoms as measured by the AOSI total score at 8 and 14 months, and ADOS total score at 24 and 36 months; MSEL = Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scale; AOSI = Autism Observation Scales for Infants; ADOS = Autism Diagnostic Observation Schedule.

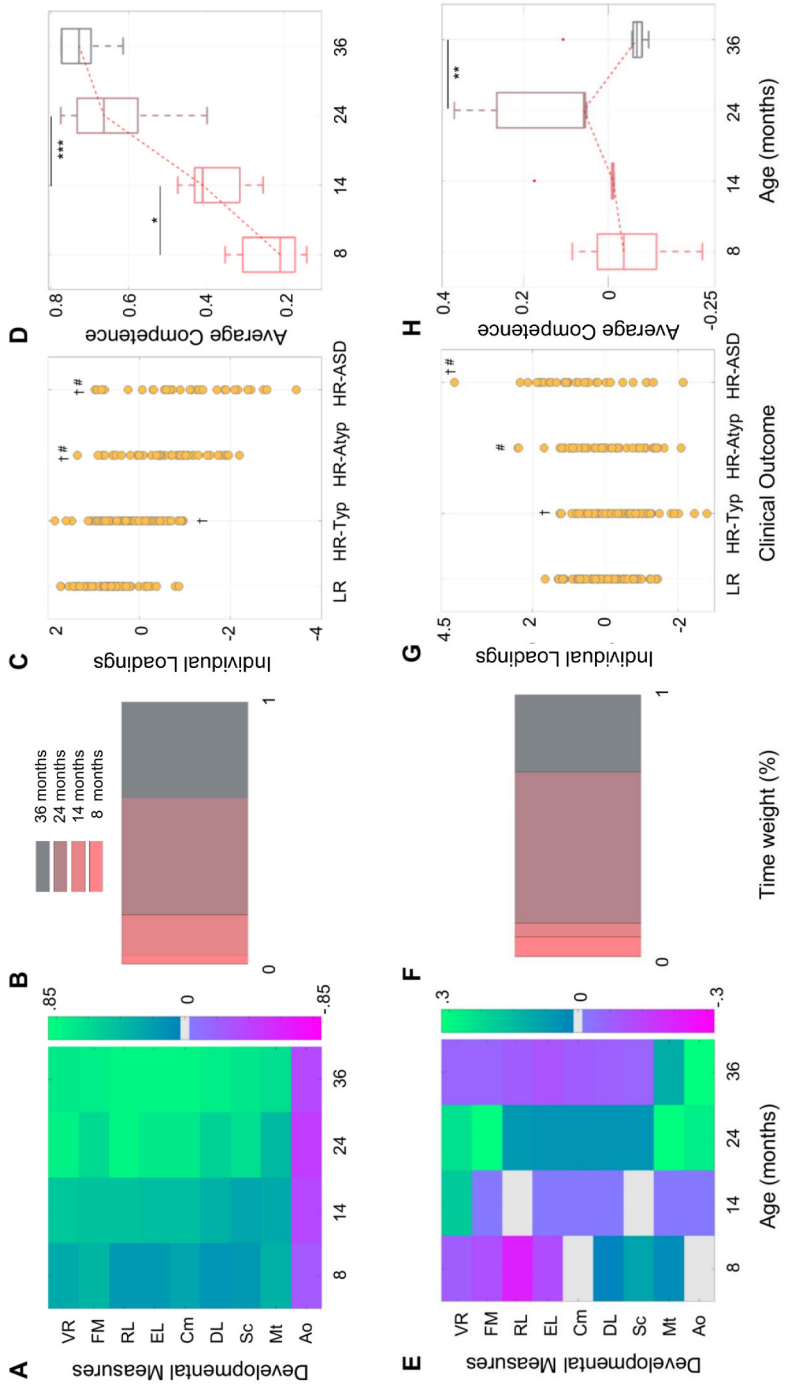


FIGURE 3. Independent components linked across development.

6

TABLE 4. Group comparisons

	IC1			IC3			IC7		
	df	t	p	df	t	p	df	t	p
LR vs.									
HR-Typical	154	4.52	$1.3{\cdot}10^{-5*}$	154	4.83	$3.3{\cdot}10^{-6*}$	131	2.01	0.046
HR-Atypical	110	9.29	$1.6{\cdot}10^{-15*}$	112	0.08	0.93	94	2.89	0.005^{*}
HR-ASD	100	7.69	$5.4{\cdot}10^{-12*}$	101	-3.45	$8.1{\cdot}10^{-4*}$	88	3.87	$2.1{\cdot}10^{-4*}$
HR-Typical vs.									
HR-Atypical	126	6.24	$6.1{\cdot}10^{-9*}$	126	-3.48	$6.9{\cdot}10^{-4*}$	107	1.13	0.26
HR-ASD	116	7.69	$5.4{\cdot}10^{-12*}$	115	-6.52	$1.9{\cdot}10^{-9*}$	101	2.21	0.03
HR-Atypical vs.									
HR-ASD	72	1.88	0.064	73	-2.59	0.01	64	0.99	0.33

Note: This table shows detailed statistics for group comparisons on the linked components significantly correlated to clinical outcome at 36 months. To correct for multiple comparisons, tests were considered to be significant (') for $p<0.05/6=0.0083$.

Abbreviations: LR = low-risk controls; HR = high-risk siblings.

DISCUSSION

This study uncovers independent neurodevelopmental processes related to clinical outcome at 36 months. We presented a data integration approach to longitudinal developmental data and early brain measures to extract intrinsic patterns of variation linked across domains. Then, we examined their relation to clinical outcome at 36 months. Contrary to supervised group comparisons, our approach exploited the power of the prospective design by moving forward in time instead of backwards. In fact, it allowed us to discover patterns in multivariate data associated with clinical outcome without any a priori knowledge on how the different measures relate to each other or to clinical outcome. In particular, looking uniquely at 8 months, a neurodevelopmental process characterized by high cognitive and adaptive competence, low levels of ASD symptoms, and higher attention to gaze shifts was associated with non-ASD outcome. Similarly, a longitudinal process characterized by increasing social-communicative and cognitive competence and low levels of ASD symptoms between 8 and 36 months, was associated with non-ASD outcome.

By integrating clinical data and ERP responses to social stimuli at 8 months, we found a single neurodevelopmental process associated with clinical outcome at 36 months. The corresponding pattern of clinical data consisted in high levels of competence and low levels of symptoms. For the corresponding ERP data the process entailed reduced attention capture but faster perceptual processing and deeper engagement to perceived gaze shifts, and reduced attention capture by visual noise. The presence of this process in individual infants, assessed through individual component loadings, showed that it was present more strongly in LR than HR-Atypical and HR-ASD infants, suggesting an association with typical development. Our previous work has shown differences in neural sensitivity to dynamic gaze at 8 months between high-risk siblings with later ASD outcome, siblings with non-ASD outcome and low-risk controls[87]. However, the ERP features differing between clinical outcome groups were not the same as those implicated in the process here. In fact, Elsabbagh et al. reported significant differences in P400 amplitude in response to gaze shifts[87], while we observed a more diffuse pattern involving P1 amplitude and latency, P400 amplitude and N290 latency. Nevertheless, this can be explained by differences in the study aims, as we extracted patterns from unlabeled data across integrated ERP measures linked to behavioural measures at the same age. Here, the profile of high behavioural and developmental scores at 8 months seems to associate with better performance on a hard task. In fact, the gaze shift condition may be more challenging for infants due to its dynamic nature involving rapid changes[87]. In particular, higher engagement to dynamic gaze at 8 months seemed to be linked to better motor and communication skills. Early sensitivity to dynamic gaze is fundamental to develop joint attention[77], which is thought to be crucial

for cognitive, language and social development[226]. Greater attention to social stimuli might provide increased opportunities for implicit social learning and the development of skills (e.g. learning words, interpreting facial expressions, predicting actions) underpinning typical development. However, the high overlap between groups in individual variation indicates that not all HR-ASD or atypical siblings were deviant on this pattern but it might rather define a subgroup.

Next, we investigated longitudinal behavioural data to identify developmental processes that might act together in the first three years of life, leading to different clinical outcome. By integrating data from standardized clinical instruments, we aimed to capture pervasiveness of ASD symptoms in multiple functional domains. We found two intrinsic developmental patterns significantly associated to clinical outcome. The first pattern indicated an increase in the average level of competence between 8 and 36 months accompanied by low levels of ASD symptoms. It occurs in a step-wise, sequential manner in which motor skills develop first, communication skills build on that and follow in development, followed in turn by social skills. The identified process was present more strongly in typical development as it explained more developmental variation in typical (LR and HR-Typical infants) than atypical (HR-Atypical and HR-ASD) infants. High-risk siblings showed significantly lower scores than LR controls while HR-Atypical did not differ from HR-ASD. However, developmental delay, poorer adaptive functioning and higher levels of ASD symptoms have been previously reported in HR non-ASD siblings[227]. Furthermore, the HR-Atypical group was more instrument-defined than clinically based and included individuals with high variability in competence and/or ASD symptoms. Among them, some individuals might develop ASD later than 36 months of age, while others might show features of the Broad Autism Phenotype[227]. The identified process is also somewhat similar to the one obtained from the multimodal analysis at 8 months, providing confidence on the identified behavioural pattern underpinning typical development. Previous studies have already shown increasing trajectories of cognitive and adaptive functioning in LR infants and HR typically developing siblings, with impairments in ASD siblings emerging in the sensorimotor domain at 6 months and moving to the social-communication domain around 12 months[35, 55]. Although the trajectory identified has already been described, the approach we used to reveal it is novel as we only considered individual-level variation across measures over time to pull apart the underlying developmental profiles. Such an approach was able to pick up this specific profile as explaining most of variance in data without any knowledge of clinical outcome. Thus, our results support and extend previous findings by showing that this longitudinal profile represents an intrinsic developmental process mostly present in low-risk and typical development. It also suggests that previously observed differences between ASD and non-ASD siblings on single behavioural measures at different time-points might actually reflect a deviation from this underlying process.

The second pattern indicated a novel profile characterized by an increase in ASD symptoms over time and a plateau in visual receptive and motor function between 24 and 36 months. This at least suggests a slower rate of gaining skills, or even stagnation over development. Contrary to the process described above, individual scores for this latter process were higher in HR-ASD siblings than LR and HR-Typical siblings, suggesting that it was present more strongly in siblings with later ASD outcome. A more far reaching interpretation is that of regression, defined as the loss of acquired skills later in development, usually between 18 and 24 months, and the later emergence of impairments typical of ASD[228]. However, developmental trajectories prior to and after the onset of regression remain unclear[229]. More recent studies have suggested that social-communication impairments were already present in infants before regression[230, 231]. Consistently, our pattern of late emerging ASD symptoms was linked to developmental impairments already at 8 months, as shown by low Mullen scores in particular for receptive language. It would be interesting to investigate whether this process could differentiate siblings who satisfied criteria for ASD already at 24 months from those who did only at 36 months. Future work should also integrate genetic data to search for genes associated with the stagnation process.

Taken together, our results highlight underlying developmental processes acting together and leading to different clinical outcome depending on their presence at the level of the individual infant. We formally investigated intrinsic processes across longitudinal behavioural and brain data, in agreement with the general consensus on the necessity for data integration to improve our understanding of the underlying mechanisms for ASD. behavioural. Our study adds to the literature by showing the actual link between multiple measures in early development, supporting the idea that early signs of ASD must be interpreted as part of larger patterns of developmental variation linked across domains and across age. The totally unsupervised approach is the strength of this study, which allowed us to pull apart different underlying processes expressing intrinsic variation in development independently from clinical categories. This approach opens up various possibilities for the investigation of the biological processes acting early in development and preceding an ASD diagnosis. In fact, future work could investigate the relation of the identified neurodevelopmental processes to different early risk factors through the integration of data from yet not incorporated modalities (e.g. MRI, fNIRS or eye-tracking data). Similarly, incorporating genetic data could aid understanding of whether a specific process is linked more to common variation or to single gene mutations. This would provide insight into trajectories of gene expression and mechanisms going from genetic risk, to neurobiological alterations and the cognitive and behavioural differences observed within ASD.

6

This study also has limitations. First, our longitudinal analysis included measures at 24 and 36 months (MSEL and ADOS) used to inform clinical outcome evaluation at 36 months, potentially biasing results. However, our main aim was to uncover underlying developmental processes and not to predict clinical outcome. The identification of these processes did not depend on clinical outcome, which was only used for post-hoc association. Thus, the identified processes were not obviously linked to clinical outcome in the first place. Nevertheless, process selection might have been biased, as shown by the fact that the identified longitudinal processes mainly explained variance at 24 and 36 months. Second, the majority in the investigated sample had a typical outcome, thus the processes identified might not capture the full variation in atypical development due to its under-representation in the sample. Third, we could not investigate the expression of neurodevelopmental processes over time as electrophysiological measures were only available at 8 months. Thus, we don't know how the identified neurodevelopmental process at 8 months develops longitudinally. However, we tested the relationship between this early process and the identified longitudinal processes through correlation analysis of the behavioural score maps. Results show a high correlation with the developmental pattern indicating an increase in the average level of competence between 8 and 36 months accompanied by low levels of ASD symptoms. Thus, the neural pattern identified from ERP data at 8 months is likely to be associated not only to high behavioural and developmental scores at 8 months, but to an increase in cognitive and adaptive functioning across development. Interestingly, the process identified at 8 months was mostly driven by ERP data (Figure 2.C), suggesting that neural measures are more informative about future clinical outcome than behavioural measures in infancy. Finally, although results show a clear significant trend relating to clinical outcome, there was substantial overlap between clinical groups. The analyses did not provide enough power to make clinical claims about proper subtyping or categorization, which are of primary interest and the focus of future research. Our study highlights the complexity of trying to separate underlying developmental processes by clinical outcome groups. Higher sample sizes, higher density of longitudinal neural data and the integration of different modalities might be needed to achieve these claims. However, although our findings do not show underlying processes specific to ASD per se, they can help shaping our view on early ASD by showing that the ASD phenotype goes beyond the limits of clinical categories set by the DSM-5 and that there is no sharp boundary between ASD and atypical development. Nevertheless, the processes identified inform on the underlying neurodevelopmental mechanisms associated with emerging ASD. Future work could link them to independent suites of genes, providing insight into the specific genetic risks for the phenotypes identified.

SUPPLEMENTAL MATERIAL

MISSING DATA

TABLE S1. Missing data. This table the number of infants attending each visit (*n/n_total*), where *n_total*=237 is the total number of infants after excluding infants who did not receive a clinical evaluation and/or an ADOS classification at 36 months; the percentage of complete data by clinical instrument at each visit; the number of subjects with missing data by clinical outcome at each visit.

Visit	Attendance *(n/n_total)*	Complete data		(%)	Subjects with missing data			(n)
		MSEL	VABS	AOSI/ADOS	LR	HR-Typical	HR-Atypical	HR-ASD
8 months	237/237	99.6	97.9	100	2	3	0	1
14 months	234/237	99.2	96.2	100	4	4	2	1
24 months	235/237	82.7	97.0	80.2	48	11	13	5
36 months	237/237	99.2	97.9	100	2	1	2	2

Abbreviations: ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; AOSI= Autism Observation Scale for Infants; ADOS= Autism Diagnostic Observation Schedule; HR = high-risk siblings; LR = low-risk controls.

Data presented in the current paper were collected as part of a large longitudinal study involving 247 infants recruited in one of two phases of longitudinal assessments (104 in Phase 1 and 143 in Phase 2). Missing data was mainly due to non-attendance to visits. N=10 infants were excluded from this study because they did not receive an ADOS (Autism Diagnostic Observation Schedule) evaluation and/or a clinical outcome evaluation at 36 months. Percentage of complete data and number of infants with missing data by clinical outcome are shown in Table S1. At 24 months, differences between infants with complete and missing data were significant on clinical outcome ($\chi^2(3)=54.2$, $p<1{\cdot}10^{-3}$). However, differences were not significant at other time-points, providing reasonable evidence for a pattern of data missing at random. To handle missing data at each time-point, we performed imputation through expectation maximization on SPSS (http://www.ibm.com/analytics/us/en/technology/spss).

Since our aim was to obtain a longitudinally complete dataset for each infant between 8 and 36 months, infants who did not attend at least one of the visits were excluded from the study (N=5, not significantly different on clinical outcome). Thus, our final sample included 232 infants (161 HR and 71 LR).

ERP TASK

Infants sat on their parents' laps at a 60 cm distance from a 40 × 29 cm computer screen.

The task was the same as in Elsabbagh et al. [87]. It was designed to assess three contrasts within the same group of infants: (1) static face irrespective of gaze direction vs. visual noise stimuli matched on spatial frequency and colour spectra; (2) static faces with direct vs. averted gaze; and (3) dynamic gaze shifts toward vs. away from the infant. The infant gaze during stimulus presentation was recorded by video camera. The visual noise stimuli were constructed from the same faces presented during the face task by randomizing the phase spectra while keeping the amplitude and colour spectra constant. Each trial block began with a static colourful fixation stimulus (subtending approximately 1.6 x 1.6 degrees of visual angle) presented for a variable duration between 800 and 1200 ms, followed by a color image of one of four female faces (subtending 21 x 14 degrees of visual angle) with gaze directed either toward or away from the infant. To ensure that infants were fixating the eye region, faces appeared in the center of the screen with the eyes on the same location as the fixation stimuli. In subsequent trials of the same block, the face remained on the screen but displayed three to six gaze shifts, alternating from directed toward to away from the infant. The visual noise stimuli were shown during approximately one third of all blocks, following the fixation stimuli as for faces. Each trial lasted for 1000ms. Trials were presented continuously for as long as the infant remained attentive.

ERP DATA ACQUISITION AND PROCESSING

EEG data were recorded using a 128 channel Hydrocel Sensor Net and EGI NetAmps 200 (gain=1000). The montage used was the same as in Elsabbagh et al. [87]. The vertex was used as a reference (Cz in the conventional 10/20 system), and data were digitized with a 500Hz sampling rate and band-pass filtered between 0.1 and 1000 Hz. Data were stored and analysed offline in EGI Netstation 5 using the same protocol as in Elsabbagh et al. [87]. Participants' overall behaviour was initially coded from videotape, and trials were retained only when infants were fixating the centre of the screen at stimulus onset, without any gaze shifts, blinking, or head movements during the segment of chosen duration (800ms in Ph1, and 1000ms in Ph2) following onset of the face stimulus or gaze shift. Data were then corrected to -200ms baseline. Following automatic artefact rejection, an experienced EEG researcher visually inspected individual trials. Rejection procedures followed established norms, including removal of segments affected by head, body or eye movement, and including those segments (identified during the video coding procedure) where the infant displayed gaze shifts or looked away from the screen during stimulus presentation. Trials were rejected when data were missing from more than 12 channels, while missing data from 12 or fewer channels were interpolated. Because of variable rates of presentation of each stimulus type, a different number of trials were included for each contrast. Stimulus-locked epochs (-200 to 800ms/1000ms peristimulus window) were averaged for the different contrasts. Visual inspection of the grand average for each condition across the three contrasts revealed characteristic task-dependent infant ERPs over occipital channel groups: P100, N290, and P400. For each contrast, the occipito-temporal channels showing the characteristic waveform were selected, avoiding any particularly noisy channels. The P100, N290, and the P400 were quantified by their amplitude and latency in response to each task.

CLINICAL OUTCOME EVALUATION

The LR group was based on risk sampling assignation and received no formal clinical diagnoses, while HR siblings received a clinical outcome evaluation at 36 months. They were subsequently grouped into siblings with ASD (HR-ASD); with atypical (non-ASD) development (HR-Atypical); and with typical development (HR-Typical).

Expert clinical researchers reviewed all available information at 24 months and 36 months and assigned clinical consensus best estimate diagnosis of ASD according to ICD-10[9] or DSM-5 criteria[8], depending on the phase of the study. The best estimate diagnoses for the two phases were reviewed for differences in categorization and considered to be similar. Among high-risk infants who did not meet criteria for ASD, a subgroup of siblings was classified as 'atypical' based on having: ADOS and/or ADI-R above ASD threshold, and/or MSEL more than 1.5 standard deviations below average on visual reception (VR) and/or receptive language (RL) and/or expressive language (EL) and/or early learning composite (ELC).

Among infants included in the final longitudinal sample, 32/161 [19.9%] high-risk siblings were categorized as HR-ASD; 43/161 [26.7%] high-risk siblings were categorized as HR-Atypical; and 86/161 [53.42%] high-risk siblings were categorized as HR-Typical. The breakdown of HR-Atypical infants based on criteria for "atypical" development is shown in Table S2.

In the subsample included in the multimodal analysis, 30/140 [21.4%] high-risk siblings were categorized as HR-ASD, 36/140 [25.7%] high-risk siblings were categorized as HR-Atypical, and 74/140 [52.9%] high-risk siblings were categorized as HR-Typical. The breakdown of HR-Atypical infants based on criteria for "atypical" development is shown in Table S2.

Table S2. HR-Atypical criteria. This table shows the number of infants among HR-Atypical siblings ($n/n_{HR\text{-}Atypical}$) who meet specific criteria for "atypical" development for the different samples included in the two separate analyses.

	MSEL criterion				ADOS criterion	ADI-R criterion	ADOS & ADI-R criterion	ADOS & MSEL criterion	ADI-R & MSEL criterion
	VR	RL	EL	ELC					
Longitudinal Analysis	9/43	14/43	9/43	14/43	30/43	6/43	4/43	7/43	1/43
Multimodal Analysis	8/36	11/36	8/36	10/36	27/36	5/36	4/36	6/36	1/36

Abbreviations: ASD= autism spectrum disorder; MSEL= Mullen Scales of Early Learning; VR= visual reception (MSEL); RL= receptive language (MSEL); EL= expressive language (MSEL); ELC = early learning composite score (MSEL); ADOS= Autism Diagnostic Observation Schedule; ADI-R = Autism Diagnostic Interview-Revised; HR = high-risk siblings.

6

7

Summary and general discussion

Autism Spectrum Disorder is a lifelong condition with onset in the first years after birth. However, a clinical diagnosis is usually established later, mostly between 4 and 5 years of age[26]. It is current practice to provide access to behavioural intervention with the direct aim to stimulate the development of social and communicative skills and reduce maladaptive behaviour and ultimate aim to improve later functional outcome[21, 22], only after a clinical diagnosis of ASD has been conferred. However, research suggests that early intervention, mainly initiated in the first year of life when the core symptoms of ASD have not emerged yet[90], might be more effective than later treatment[91, 92, 232, 233]. To enable early targeted intervention, it is important to identify the infants who need intervention early in life. Previous findings on group differences tell us little if anything about the individual infant as there can be significant overlap between groups in individual variation. Thus, this thesis focused on translating group differences to prediction of ASD outcome at early age at the level of the individual infant. In addition to early detection, prediction at the individual level provides more insight into the cognitive and neural mechanisms underlying ASD development through the identification of potential biomarkers for the disorder. Previous studies have already attempted to predict individual ASD outcome in the first 2 years of life; however, the high clinical, biological and etiological heterogeneity of ASD[94] hampers early detection at an individual level. To overcome this limitation, I combined different types of data in multivariate models for prediction of later clinical outcome by using machine learning techniques for supervised learning. Another limitation might come from the supervised case-control comparison approach based on clinical labels defined at the time of diagnosis, which do not consider the heterogeneity of ASD[121] and might not even have meaning earlier in development. Thus, in the second part of the thesis I used unsupervised learning to capture unknown structure in data and investigate the relationship with current categorical outcomes. In this final chapter, I will first summarise findings from the studies presented in this thesis and discuss them in relation to the current literature. Next, I will discuss the limitations of this work and provide directions for future research, and I will conclude with final remarks.

SUMMARY OF RESULTS

The first part of this thesis (chapters 2 to 4) focuses on the investigation of early signs of and precursors to ASD in the first two years of life through supervised classification techniques. In the second part (chapters 5 and 6), we move to unsupervised approaches to stratify ASD heterogeneity in terms of subgroups of infants and the cognitive and brain mechanisms involved.

In Chapter 2, I used a support vector machine classifier to predict ASD outcome at the individual level among HR siblings using measures of developmental level, adaptive functioning and early ASD symptoms collected at 8 and 14 months. Furthermore, I characterised group trajectories of developmental level and adaptive functioning between 8 and 36 months in low-risk controls (LR), and high-risk siblings with an outcome of typical development (HR-Typical), atypical development (HR-Atypical), or ASD (HR-ASD) at 36 months. At the group level, LR and HR-Typical showed higher developmental level and functioning over time than HR-Atypical and HR-ASD. At the individual level, prediction of ASD among HR siblings was possible with 71% AUC using measures of daily living skills at 14 months. The classifier had a much lower positive predictive value (PPV=29%) than negative predictive value (NPV=92%), which means that it was more accurate at predicting infants who will not develop ASD. This can be clinically useful by predicting typical development as it allows the identification of children who don't need to be intently surveilled.

In Chapter 3, I investigated temperament as an early risk marker for ASD by examining parent-reported questionnaires for HR and LR siblings at 8, 14, and 24 months. Similarly to Chapter 2, I characterised developmental trajectories at the group level and used a support vector machine classifier to predict individual ASD outcome at 36 months among HR siblings. At the group level, I observed significant differences between clinical outcome groups, with more atypical temperament for HR-ASD, followed by HR-Atypical, HR-Typical, and LR siblings. At the individual level, prediction of ASD outcome was possible with 71% AUC using measures of effortful control at 24 months but positive predictive value was low (PPV=29%). However, a negative predictive value of 96% indicated that while siblings with later ASD could not be identified accurately, HR infants without ASD could. This suggests that temperament may not facilitate early identification of ASD individually, but it may help identify HR infants who do not develop ASD.

In Chapter 4, I investigated group and the individual level associations of neural sensitivity to visual noise and faces with static gaze and dynamic gaze shifts in 8-month-old infants with ASD outcome at 36 months. At the individual level, I achieved a classification accuracy

of approximately 80% AUC, with 81% PPV and 71% NPV, using speed and amplitude of response at early and later stages of face processing. Such a pattern converged with results at group level to indicate that infants developing ASD show broad difficulties reflecting general alteration in neural processing of faces already in infancy.

In Chapter 5, I used parallel growth mixture modelling to identify distinct classes among infants at high and low familial risk for ASD based on early development of adaptive functioning between 8 and 36 months. I observed three distinct trajectories: initially high but then decreasing scores across all VABS scales (8.3%); relatively stable trajectories around age-appropriate norms (73.8%); and initially low but then increasing trajectories and reaching a stable average level in all scales by age 2 (17.9%). Decreasing trajectory of adaptive behaviour had a significantly increased risk [OR=4.40 (CI: 1.90; 12.98)] for an ASD outcome compared to the other trajectories, and higher parent-reported symptoms in the social, communication and repetitive behaviour domains at 36 months. Furthermore, I observed the emergence of a pattern showing discrepancy over development between adaptive behaviour and cognitive level. In particular, the improving class in adaptive behaviour showed stable trajectories of cognitive development around average scores.

Finally, in Chapter 6 I uncovered processes affecting multiple developmental domains by using unsupervised learning techniques and selected the processes significantly associated with clinical classification at 36 months. In particular, I examined developmental level, adaptive functioning and early ASD behaviours between 8 and 36 months, and neural sensitivity to eye-gaze at 8 months in infants at high and low familial risk for ASD. I identified two independent, longitudinal processes and one early stage process having coherent effects across multiple domains of development. The early process at 8 months was associated with non-ASD outcome and it was characterised by high functioning and low levels of ASD behaviours linked to reduced attention capture but increased engagement with gaze shifts, and reduced engagement with non-social stimuli. Then, one longitudinal process was associated with non-ASD outcome and characterised by increasing cognitive and adaptive functioning and low levels of ASD behaviours, while the other was associated with ASD outcome and indicated a stagnation in cognitive functioning at 24 months.

GENERAL DISCUSSION

This thesis addressed two main issues in the investigation of ASD development in infancy: prediction of later ASD diagnosis at the level of the individual infant, and understanding of heterogeneity in ASD manifestations by extending our knowledge of cognitive and neural mechanisms underlying the development of autism. The common point was the transition from the level of the group to the level of the individual infant.

BRAIN DATA OUTPERFORM BEHAVIOURAL DATA IN INDIVIDUAL PREDICTION OF ASD IN THE FIRST YEAR OF LIFE

This thesis shows that early detection of ASD at the level of the individual infant is more difficult to achieve from measures of behaviour than brain function, and data integration of behavioural and brain data does not really improve prediction compared to brain data alone.

There is a general consensus in the field that the defining behavioural features of ASD are not present in the first year of life but begin to emerge around 12 months and consolidate between 18 and 36 months[212, 219]. However, this pre-symptomatic period is characterised by sensorimotor[35, 36, 40, 55] and visual attention[38, 220-222, 234] atypicalities in those infants with later ASD outcome. Few studies so far have tested individual-level prediction from behavioural data in the first 2 years of life. In particular, parental concerns at 6 months showed an AUC of approximately 68%, which is not significantly different from chance level, for the prediction of high-risk siblings with later ASD from high-risk siblings with other developmental disabilities. At 12 months, the same measures showed a higher AUC of 74% but only 58% PPV[235]. Similarly, behavioural profiles built on the ADOS scores at 18 months could predict later ASD outcome with 77% overall accuracy in a validation sample, but only 50% PPV[114]. In this thesis (Chapter 2,[55]), prediction of ASD outcome using integrated, standardized clinical measures from multiple functional domains and multiple time points was at chance level at 8 months, but at moderate and above chance level accuracy (AUC=71.3%) using measures of daily living skills at 14 months, although PPV was only 29%. The same applies to temperamental data between 8 and 24 months (Chapter 3), since group differences in temperament traits detectable in infancy did not translate into an acceptably accurate prediction of ASD in the individual infant. In fact, effortful control at 24 months had the highest predictive power (AUC=71%) and a high negative predictive value (NPV=96%), but low positive predictive value (PPV=29%) for individual prediction of ASD and atypical development. While it is possible that temperament is masked by early ASD symptoms and cannot accurately predict positive cases, good effortful control seems to be useful to allay concerns for later ASD outcome. Overall, results suggest that, although

behavioural differences between groups of infant siblings with later ASD diagnosis and unaffected siblings or low risk controls can be detected in the first year of life[34, 35, 55], behavioural markers alone are not sufficiently specific or sensitive to predict later ASD outcome at an individual level. Behavioural data are probably too noisy and not specific to ASD in this prodromal phase, when the characteristic symptoms of the disorder have not fully emerged yet, thus it likely cannot be distinguished from other developmental disorders. The same happens, for instance, when getting the flu: the prodromal signs may be subtle and varied (fatigue, sore throat, headache, sleep problems, coughing, runny or stuffy nose, concentration problems), but symptoms are more similar across patients once the disease reaches its full manifestation. In fact, integrated data from standardized clinical instruments allowed us to predict broader atypical development (*HR-ASD + HR-Atypical vs HR-Typical*, Chapter 2) already at 8 months with 69% AUC, significantly different from chance level, and 63% PPV using a combination of motor and communication scores; while predictive power increased to 71% AUC with 62% PPV at 14 months using a combination of scores from adaptive skills and early ASD symptoms[55]. Nevertheless, results from the studies reported in Chapters 2 and 3 show high negative predictive value for classification of ASD vs. non-ASD siblings using respectively measures of daily living skills at 14 months of age and measures of effortful control at 24 months. This might indicate that the path to typical development is more coherent as prediction of typical development using early behavioural and temperamental data was possible with higher accuracy compared to ASD outcome. Thus, good daily living skills and good effortful control early in development seem to carry on and can be useful indicators to allay concerns for later ASD outcome.

Although still not useful as clinical diagnostic markers at an individual level, early behavioural differences in infants with later ASD suggest that there might be preceding or co-occurring alterations of neural features. Previous studies have shown that alterations in brain structure[112, 212, 223, 224] and function[86, 87, 111] are present during the prodromal phase of ASD and can predict later ASD outcome at an individual level[219]. In particular, excessive extra-axial cerebrospinal fluid (EA-CSF) at 6 months predicted which infants would receive an ASD diagnosis at 24 months among high and low risk infants with 72% AUC, 80% sensitivity and 67% specificity[224]. Similarly, prediction of ASD outcome at 24 months from a combination of MRI functional connectivity data in 6-month-old high-risk siblings was possible with 97% accuracy and 100% PPV. These differences in brain structure and function were evident and discriminative for ASD before the emergence of the core behavioural features of the disorder, and approximately at the same time when behavioural differences in the sensorimotor domain become detectable[33, 35, 46]. In fact, excessive EA-CSF was significantly correlated with poorer motor skills at 6 months[224]. A combination of measures of cortical surface area, cortical thickness and intracranial volume between 6 and 12 months

predicted ASD outcome at 24 months with 94% accuracy and 81% PPV among high-risk siblings[112]. These cortical changes, together with atypical white matter structure[210], are concurrent with the emergence of sensory and attentional deficits associated with ASD in its prodromal phase[35, 219, 222].

A conceptual framework for ASD

Supported by these findings described above, a recent conceptual framework has been proposed for the development of ASD[212]. According to this framework, an early proliferation of neuroprogenitor cells leads to the hyper-expansion of cortical surface area, as measured at 6 months, which in its turn leads to the sensorimotor and visual attention deficits observed in the ASD prodrome. These early alterations in basic attentional and sensorimotor function affect experience-dependent neuronal development[236], in particular the refinement of selected neural circuits through pruning and apoptosis, ultimately leading to the social deficits characteristic of ASD and consolidated between 18 and 36 months of age, when brain volume overgrowth and disrupted connectivity have also been reported. Figure 1 illustrates a schematic representation of the conceptual framework for ASD and how the classification studies reported in this thesis relate to it.

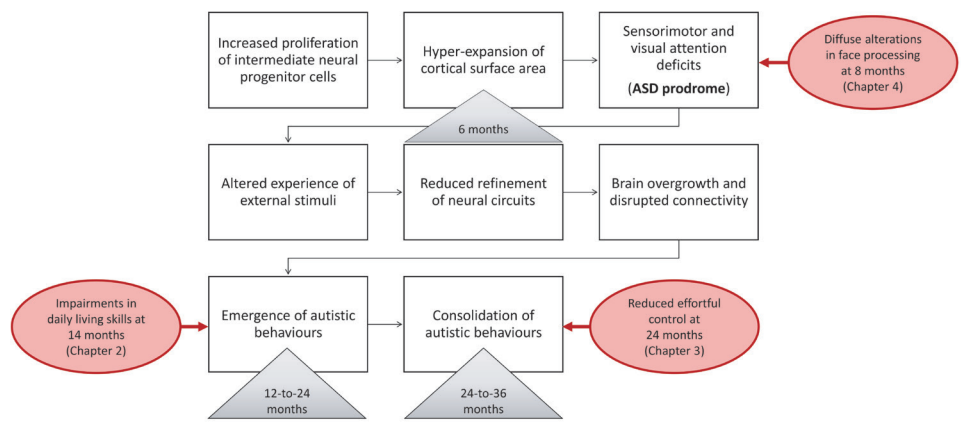


FIGURE 1: a conceptual framework for ASD. This figure illustrates a schematic representation of the conceptual framework described by Piven et al.[212] for early development of ASD, and how our findings on early predictors of ASD relate to it.

In line with the early emerging ASD behavioural phenotype described in this framework, our ERP study (Chapter 4) illustrates early sensory and attentional deficits at 8 months in high-risk siblings with later ASD outcome. In particular, group-level and individual-level

analyses on neurophysiological measures of face processing at 8 months converged to indicate a diffuse pattern of alterations of face processing which allowed prediction of ASD outcome with approximately 81% AUC and 81% PPV already at 8 months (Chapter 4). This pattern was reflected in both early sensory stages and later, higher-level stages of information processing of dynamic gaze and faces as opposed to visual noise. This points to a general alteration in neural processing of faces as an indicator of later ASD. Altered experience due to diffuse deficits in processing may indeed result in downstream changes leading to the social deficits that are more traditionally associated with the ASD condition. Thus, early measures of neural processing, as allowed by EEG and ERP studies, represent a useful instrument for early detection of ASD in its prodromal phase, while understanding the impact of altered attentional and sensorimotor function in infancy may also provide useful insight for early interventions[91]. However, very few studies so far have used EEG/ERP to directly compare face processing between ASD and non-ASD outcome groups in infancy[86, 87]. Our study (Chapter 4) represents the first ERP study using a machine learning approach to predict individual ASD clinical diagnosis in HR infants at 36 months, while one previous study on an overlapping dataset used logistic regression on ERP responses to gaze shifts at 7 and 14 months to predict ASD clinical diagnosis at 7 years of age[177]. Compared to other studies using functional and structural brain imaging data to predict ASD outcome in the first year of life[111, 112], we extended results to a more cost-effective, mobile, and infant-friendly neuroimaging technology (ERPs). This would be more feasible in clinical practice than use of MRI technology, thus increasing accessibility of early screening. Furthermore, although previous studies have shown promising results for early prediction of ASD diagnosis at 24 months, ample research has demonstrated that some children will only manifest the diagnostic features of ASD later in life[230, 237]. The work presented in this thesis instead focused on prediction of the best estimate ASD diagnosis at 36 months, which has previously shown to have a good stability up to 10 years[238].

Does integration of data from different domains improve prediction?

One of my hypotheses was that the integration of data from different functional domains and different time points would improve predictive value for ASD at an individual level in the first year of life. However, this thesis shows that data integration does not improve classification of ASD from behavioural and temperamental data at 8 months, which remained at chance level (Chapters 2 and 3), while it improves prediction of broader atypical outcome from behavioural data (Chapter 2). Furthermore, the integration of neural (ERP) and behavioural (MSEL, VABS and AOSI) data showed that behavioural data do not add value for individual prediction of ASD at 8 months (*Additional material*, Chapter 4). In fact, behavioural scores were not selected among the relevant features for prediction of ASD by the genetic algorithm. These findings are consistent with findings from a previous study, showing that neurocognitive responses to gaze shifts at 7 months improve predictive power for ASD outcome in toddlerhood at an individual level over and above measures of ASD symptoms severity at 7 months[177]. This leads back to the heterogeneity of the behavioural phenotype in the ASD prodrome, when behavioural atypicalities are subtle, not sufficiently sensitive to ASD at an individual level and/or not specific to ASD. In line with it, the combination of behavioural data at 8 and 14 months did not improve prediction compared to data at 14 months, when behavioural signs are stronger and provide higher predictive value (Chapter 2). It is, in fact, in the second year of life that the defining behavioural features of ASD begin to unfold. On the other hand, ERP data alone at 8 months provided good predictive value for ASD at an individual level likely because measuring the early sensory and attentional alterations that are more commonly described as part of the early emerging ASD phenotype[212, 219].

EXPLORATION OF ASD HETEROGENEITY TO MOVE FORWARD INSTEAD OF LOOKING BACKWARD

The first part of this thesis highlighted the high heterogeneity within (i.e. in terms of changes over development) and across individuals in the ASD category. The overall large overlap of categorical outcome groups (typical versus atypical versus ASD) in individual behavioural variation hampers prediction of clinical outcome as defined by current clinical labels. Although pattern identification methods based on machine learning techniques can improve early detection of ASD at an individual level compared to group-based analyses, supervised classification is still centred on the concept of ASD as a single predefined category. This pointed to the necessity to further investigate individual heterogeneity in ASD to understand its manifestations and improve stratification so that a more accurate prediction would be possible early in development. Thus, in Chapters 5 and 6 I moved from traditional retrospective analyses based on comparisons of outcome groups defined around three years of age to focus on prospective analyses looking at structure in the data early in infancy, irrespective of predefined categories.

ASD is characterised by high clinical heterogeneity expressed as considerable variability across individuals in onset, cognitive function, language skills, symptoms profiles, severity, and psychiatric and neurological comorbidities. This variability in manifestations reflects the high etiological heterogeneity of the disorder[30, 94, 98, 99]. Causative contributions to ASD include environmental and medical risk factors (e.g. prenatal drug exposure, preterm birth or congenital infection), and hundreds of genetic variants impacting specific molecular pathways and regulating fundamental processing of early brain development (e.g. cortical organization, excitation/inhibition balance and connectivity), ultimately leading to distinct clinical presentations[100]. This heterogeneity makes it difficult to understand, to detect and

to treat ASD as a single disorder at the level of the individual infant. While it seems a logic consequence to abandon the unitary ASD diagnostic label and investigate the complexity behind it, the case-control comparison methodology is still the main approach in clinical and translational research. Although useful on a practical level for providing a diagnostic label, the idea of ASD as a discrete, separate entity can distort the investigation of the underlying mechanisms and early development of ASD. Based on previous findings supporting the common etiology of autistic traits in individuals with ASD and the general population[239, 240], ASD would be better described at an individual level as reaching extreme values in a spectrum of continuously varying traits in the general population[241] rather than as a well-defined, separate category.

Major initiatives in child psychiatry[121] have started to focus on parsing the heterogeneity of ASD and identifying biological subtypes so to improve early detection and enable mechanistically targeted interventions based on the specific pathophysiology of the identified subgroup. Stratification of ASD heterogeneity could be done on multiple levels, from etiology to neural systems, cognition, behaviour and developmental patterns[122]; the aim is however the same: the identification of distinct, replicable subgroups within the ASD population, naturally organized in more robust differences than between ASD as a single group compared to a matched control group. Data-driven methods are particularly advantageous to this purpose due to the absence of hypotheses for the inference of structure in unlabelled data, making them particularly useful when there is no a-priori knowledge on the sample subgroups[167]. Among them, normative modelling has been used in neuroimaging[242] to evaluate deviance from normative ranges at an individual level by mapping the full range of population variation and considering symptoms in patients as deviations from the normative pattern[243]. There are well known limitations in the different ways partitioning can be obtained based on the chosen measures and algorithm, the size of some subgroups that can be very small, the relatively arbitrary choice of number of clusters, and stability of partitions over time. Nevertheless, clustering methods remain the most commonly chosen methods to fractionate clinical groups at a particular time point[122, 244].

Chronogeneity of ASD

Of note, heterogeneity of ASD is also evident in developmental trajectories, for which the concept of "chronogeneity" has been introduced to indicate group and individual variability of development over time and the possibility for individual infants to deviate from a group trajectory[97]. This highlights the necessity for dense longitudinal assessments and longitudinal data analyses approaches to characterise the different developmental trajectories leading to and within the ASD outcome. Similar to clustering, latent class analysis[245] is a "model-based clustering" approach which has been employed for the identification of subgroups in clinical cohorts[246-249] and assumes that a "latent" process underlies the structure of the observed data, in our particular case ASD development. Based on finite mixture modelling, latent class analysis finds the probabilistic model fitting the data distribution and then assesses the probability of membership to a certain class case-by-case[250]. An advantage over clustering techniques is the possibility to assess model fitting as an evaluation metric for goodness of data modelling. Latent class analysis can be adopted to identify subgroups in clinical cohorts based on cross-sectional data but also longitudinal profiles. In the latter case, it is called latent growth curve analysis[123] and it has been commonly used to identify classes in developmental trajectories within ASD and high-risk cohorts[54, 198]. In this thesis, I chose an extension of latent growth curve analysis, growth mixture modeling[125], which allows for within-class variation to obtain a more realistic representation of heterogeneity in longitudinal development of adaptive behaviour in infants at high and low familial risk for ASD (Chapter 5).

Previous studies have used longitudinal modelling to explore heterogeneity in children already diagnosed with ASD. While analyses on ASD symptoms led to mixed findings[4, 191, 251-253], findings on adaptive behaviour in pre-school and school-aged children with ASD were more coherent[191-195], showing trajectories with increasing, stable and decreasing scores over time. This suggests that developmental trajectories of adaptive skills may actually serve to define subgroups in ASD phenotype. Furthermore, the investigation of early development of adaptive skills might be critical to predict later functioning in different environments of everyday life and then enable early targeted intervention aimed at improving later functional outcome through learning of adaptive skills. Thus, I focused on adaptive behaviour for stratification of heterogeneity in developmental trajectories preceding an ASD diagnosis. The study shown in this thesis (Chapter 5) represents one of the few studies investigating latent classes of infants at high and low risk for ASD without a formal diagnosis yet[54, 198, 251, 254]. Overall, children referred for possible ASD between 1 and 4 years old seem to show five different trajectories of ASD symptoms: three stable trajectories at different severity levels (mild, moderate and severe), and 2 improving trajectories (one from severe to moderate scores, and one from moderate to mild scores)[251]. Consistently, five classes have been identified based on ASD symptoms severity and cognitive skills through the first 3 years of life[254]: two classes with low symptoms severity and respectively high and average cognitive skills; one class with high symptoms severity but high levels of verbal and nonverbal functioning; one class with low symptoms severity and low cognitive levels, especially in the verbal domain; and one final class with high symptoms severity and low cognitive level. One recent study from Sacrey et al.[198] has investigated latent trajectories of adaptive functioning based on the adaptive composite

score from the Vineland in a high-risk cohort of 566 infants between 12 and 36 months. Results showed one class with average scores at 12 months and a declining trajectory, one class with a slightly declining trajectory, and one class with higher scores and a stable trajectory. Here, I examined younger ages and found three main latent classes based on adaptive functioning between 8 and 36 months: one with decreasing scores on all scales of adaptive behaviour; one with overall stable scores around age-appropriate norms; and one with increasing scores from below age-appropriate norms before age 2 and stable average scores afterwards. Looking at trajectories from the second year of life, our findings are consistent with those from Sacrey et al.[198]; however, the inclusion of an earlier time point allowed us to examine the high variability of adaptive functioning in the first year of life. This added particular value to our findings by showing that children with ASD outcome might not simply follow a trajectory of progressive impairment in adaptive skills, but some of them might present even stronger skills in the first year of life compared to other subgroups of infants.

Checking trajectories of cognitive development for the identified classes, there was a discrepancy between cognitive and adaptive functioning in the sense that trajectories of cognitive development were only two in practice: decreasing scores for the class with decreasing adaptive functioning, and stable average scores for the classes with average and increasing adaptive functioning. Evidence from previous studies on adults with ASD supports the discrepancy between cognitive ability and adaptive functioning, suggesting that this discrepancy might also increase with age[199, 200]. However, the two main levels of cognitive development observed in our cohort are in agreement with the study reported above[254]. Furthermore, classes did not differ in clinically observed symptoms severity but there were significantly higher parent-reported symptoms in the social, communication and repetitive behaviours domains, and a significantly higher risk for ASD outcome for infants with declining adaptive skills. The lack of correspondence between adaptive skills and clinical observations of early symptoms emerging from our study is consistent with previous findings in older children with ASD[191], showing a complex relationship between the presence of symptoms and the development of everyday functioning. However, this split might be also due to the parent-reported nature of adaptive behaviour scores used to identify classes of infants, as differences between classes were significant in parent-reported symptoms but not clinically observed symptoms.

ASD is a lifelong condition characterised by persisting impairments in language, social skills and daily living functioning[18], and the investigation of early development of adaptive skills can be critical to predict later functional outcome, independently from the observed symptomatology, which can even be masked by learned strategies to cope

with environmental demands[8]. In particular, our study allowed to identify subgroups that might be more relevant target groups for intervention aimed at improving functional outcome. Interventions need to focus on both adaptive skills and early symptoms because improvement in symptoms severity does not ensure improvement in everyday functioning[191]. Furthermore, the choice to investigate subgroups based on developmental trajectories linked across the different domains of adaptive behaviour, instead of analysing the composite measure of adaptive functioning, provides a more detailed insight into the strengths and weaknesses for individuals in a specific group, and might thus help better targeting personalized interventions early in infancy. In fact, given the importance of adaptive behaviour for everyday functioning in different environments, early access to personalized interventions like the Early Start Denver Model[255], whose effects have shown to generalize to everyday life and prevent infants from falling further behind age-appropriate norms in adaptive behaviour[256], might be crucial to improve later functional outcome. This is particularly true for infants in the declining class, showing a higher risk for ASD development in toddlerhood but also higher adaptive skills in the first year of life, on which early intervention can build to prevent plateauing of skills and to improve later functional outcome.

Heterogeneity as useful information to discover the underlying mechanisms of ASD development

Given the heterogeneity and chronogeneity of ASD, it becomes important to integrate information from multiple concurrent and longitudinal data to decompose this variability[100] and understand the complexity of ASD development. In this thesis, I chose a novel unsupervised approach (FLICA) for the extraction of intrinsic patterns in multivariate data, focusing more on the identification of underlying processes rather than subgroups. Potentially, the next step is to employ individual scores on these processes for clustering to stratify the sample under investigation. This approach allows for a better understanding of the underlying mechanisms leading to different subgroups in phenotype by investigating their effects across multiple behavioural and developmental domains over time, and across modalities. In particular, I found a longitudinal pattern of impaired development of social-communicative and cognitive competence and high levels of symptoms associated with ASD outcome, and captured what seemed to be a stagnation or even regression process across domains of cognitive and adaptive functioning. Looking uniquely at 8 months instead, the identified pattern associated to non-ASD outcome suggested that being very typical in cognitive and functional development early on is associated with a better processing on a hard visual task such as dynamic gaze shifts, showing reduced attention capture but faster perceptual processing and deeper engagement with gaze shifts. In addition to previous studies, Chapter 6 introduces a novel approach for prospective analysis as opposed to

the more traditional retrospective investigation of early differences between categories defined by ASD outcome. This allows to understand the different emerging patterns of development and how they lead to specific outcomes by looking at structure in the data, rather than starting from the clinical categories defined after diagnosis later in life, which might not even exist earlier in development. The identified patterns might then be the key to improve our understanding of individual heterogeneity and allow stratification into more homogeneous and predictable subgroups.

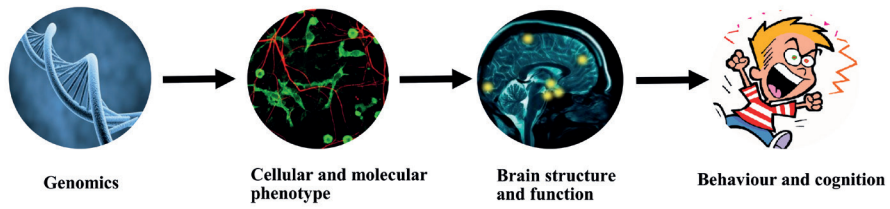


FIGURE 2: a conceptual framework from genes to behaviour. This figure illustrates a schematic representation of a conceptual framework according to which genes contribute to behaviour and cognition in ASD via their effects on cellular and molecular phenotype, affecting in its turn brain development.

The main strength of the FLICA approach lies in its potential to integrate different kinds of data within the same individual, like genetic, neuroimaging, neurocognitive and behavioural data. This would allow to investigate links between genetic, biological and behavioural domains, providing insight into the mechanisms going from genetic risk, to neurobiological alterations and the cognitive and behavioural differences observed within ASD (Figure 2). Stratification of ASD based on these underlying processes would lead to more biologically meaningful subgroups, that might be a better target for early intervention, compared to stratification based on behavioural differences only[257]. In fact, phenotypic variability within ASD and its overlap with other neurodevelopmental disorders, like ADHD, could be due to different genetic alterations affecting synaptic development through neurogenesis, dendritic projections, and neuronal migration, and resulting in abnormal development across domains and across disorders[212]. Penetrance of these alterations may depend on their total amount, the timing of consequential alterations in synaptic development and circuitry formation, but also on the influence of gender, epigenetics and environmental effects[257]. In this scenario, FLICA would allow us to integrate information from different domains and different modalities considering individual heterogeneity no more as noise but as useful information to discover the different pathophysiological mechanisms leading to ASD[258, 259].

LIMITATIONS AND FUTURE DIRECTIONS

See Box 7.1 for a brief overview of suggested future directions emerging from this thesis.

BOX 7.1: Future directions

- More research is needed to test generalizability of findings on early detection of ASD to real-life clinical settings, for instance through large, multi-site, prospective community-based studies.
- Future research should investigate challenges faced by families of children diagnosed with ASD to provide adequate support.
- More research is needed to make effective interventions accessible at the age when early detection is achieved.
- Specificity and sensitivity of early markers for ASD need to be investigated when presented with comorbidities before being translated into clinical practice.
- Future research should investigate early manifestations of ASD according to a specific genetic etiology.
- Future studies should stratify high-risk siblings based on genetic/cognitive/clinical characteristics of the older sibling with ASD.
- Future studies should stratify high-risk siblings based on genetic/cognitive/clinical characteristics of the parents.
- New high-risk cohorts, such as preterm infant cohorts, need to be investigated to assess all potential risk factors for ASD.
- Future research should investigate early manifestations of ASD based on birth order.

CLINICAL IMPLICATIONS

Early detection of ASD, before the core symptoms emerge, has important clinical and ethical implications. Although brain data provide the highest predictive value for ASD in infancy, it is unlikely that the use of biological predictors will replace expert clinical diagnosis. Rather, machine learning classifiers may serve as a supplementary screening tool for clinicians to indicate the risk that an individual infant has for developing ASD[89], providing a richer endophenotype than behavioural evaluation alone. However, the effects of early prediction on families and society need to be taken in consideration. Positive genetic testing in other medical areas has shown limited or no adverse psychological effects on families[260, 261], and in some cases early screening has proven to improve family coping[262]. Nevertheless, early detection may affect societal perception and a false positive prediction might cause children and their families to experience stigma or devaluation[263]. On the other hand, false negative prediction can delay access to intervention and services

needed by children with ASD. Thus, additional research needs to focus on the challenges faced by families managing the child's condition to provide adequate support[264, 265]. Furthermore, early detection becomes clinically useful only if early access to effective interventions is possible at the age of detection. Thus, future work should also focus on the quality and accessibility of early intervention programs.

A main contributor to ASD heterogeneity that plays a fundamental role in clinical practice but was not taken into account in this thesis is the extensive comorbidity with other psychiatric conditions[266-268]. In fact, 70% of children with ASD have been reported to show at least one comorbid disorder, and 41% of them have shown two or more[269]. Among the different comorbidities, we can list ADHD[270], intellectual disability[271], anxiety disorders[272], Oppositional Defiant Disorder[269], depression[273], gastrointestinal disorders[274] and epilepsy[275]. These comorbid conditions could work as mediating or moderating factors for the emergence of symptoms[276, 277], however, few studies have attempted to investigate the co-occurrence patterns of comorbidities in ASD[278]. Future work on early detection and stratification of ASD should carefully take into consideration comorbidities by integrating clinical and biological measures assessing signs and symptoms for different conditions, improving our understanding of the underlying biological mechanisms and precursors of this broader range of conditions. In particular, to validate ASD biomarkers it might be crucial to investigate how they vary with the presence of comorbidities both on a dimensional and a categorical level, and examine the specificity and sensitivity of these markers to ASD when presented with multiple disorders. In fact, a differential diagnosis is often required in clinical practice, rather than classification of patients from controls, and the same individuals may be part of the different diagnostic classes at the same time due to comorbidities. Thus, future research should accommodate this before being translated into clinical practice, for instance through multi-label classification[279] or multi-task learning[280].

Other factors to carefully consider in predictive modelling are sample size and generalizability. Many studies suffer from relatively small sample sizes and estimation of generalizability within a single study[281], while accuracy estimates seem to decrease with sample size[106, 281]. Generalizability of predictive models must be tested through proper validation on independent cohorts, different laboratories and different testing conditions before being considered clinically useful[106]. The studies presented in this thesis included relatively large sample sizes and generalizability was tested through cross-validation; however, external validation needs to be performed to test generalizability of our findings to independent cohorts. Furthermore, research mainly focuses on samples that typically are not representative of the general population, limiting generalizability of findings to real-life clinical settings. Future research should focus on large-scale community cohorts and use multi-site consortium datasets to improve generalizability for clinical practice.

BEYOND THE HIGH-RISK DESIGN

Although valuable for identifying early risk markers for ASD, the high-risk design suffers from several limitations. First, high-risk infant cohorts are heterogeneous on genetic profiles as recruitment is not specific to the genetic risk factors or more generally the etiology of ASD of an individual infant. Future studies should investigate early manifestations of ASD according to a specific genetic etiology, informing on the associated penetrance of neurobehavioural manifestations and symptomatology, and compare between syndromic and non-syndromic forms of ASD[13, 282], and between syndromes. This strategy will allow to investigate whether there is a final common biological pathway to ASD, or there are distinct mechanisms underlying the different manifestations of ASD. This will require cohorts defined by genotype[283] and followed-up longitudinally to identify the altered biological pathway linked to that specific genotype and leading to a certain phenotype later in life. Once identified the dysregulated pathways specific for that genetic group, personalized intervention can be enabled in the first year of life. In particular, monogenic syndromes such as tuberous sclerosis complex and copy number variants (e.g. 22q11.2 deletion syndrome), which can be diagnosed at a prenatal stage and have an increased incidence of ASD to 60%, provide a unique opportunity to investigate the underlying mechanisms specific to that syndromic form of ASD in its prodromal stage. A challenge in these cases would be reduced sample sizes; however, the identified cohorts would likely be less heterogeneous and improve accuracy of diagnostic biomarkers at early age. Second, stratification of high-risk infants is usually not at all performed based on characteristics of the older sibling with ASD. It would be of great interest to investigate whether and how gender, cognitive level, presence of a specific genetic variation or symptoms profile of the older sibling with ASD affect early development of the younger sibling leading to different developmental outcomes, including ASD or other developmental disorders. The same applies to the parents' genetics and possible conditions. A recent study has shown, in fact, an indirect effect of non-transmitted genetic variants of the parents on the fitness of the child, defined as "genetic nurture"[284]. This effect is mediated by the environment that the parents create for their child, and it might actually be the same for other member of the family like an older sibling with ASD. Thus, future studies should take into consideration family history and genetic variations to have a better understanding of the child development in relation to the environment. Third, early ASD development in first-borns cannot be investigated by design in prospective high-risk studies. In fact, infants are recruited based on having at least an older sibling already diagnosed with ASD. However, previous studies have shown that first-born males are overrepresented in the ASD population[285-287]. Future research might employ community-based cohorts to compare early manifestations of ASD based

on birth order. Finally, future studies should extend the investigated cohorts to new risk groups, such as preterm babies[288], to investigate all potential risk factors for the disorder and early manifestations associated with that specific risk.

FINAL REMARKS

This thesis indicates a diffuse pattern of neural responses to faces and visual noise at 8 months as a possible precursor of ASD outcome at 36 months, showing good predictive accuracy in the first year of life when behavioural signs are still not sensitive and specific enough for prediction at an individual level. This points to a general alteration in neural processing of faces as an indicator of later ASD. Furthermore, data-driven analyses allowed us to exploit the power of the prospective design at its full to identify subgroups among high-risk siblings and low-risk controls, and to uncover underlying processes acting together early in development and associated to ASD outcome. Key findings are reported in Box 7.2. The identification of homogeneous subgroups within ASD and the development and validation of biomarkers remain the key challenges for future research. The goal is to combine predictive modelling with existing clinical expertise in behavioural assessments to obtain an earlier and more reliable detection of ASD, and to determine the best treatment and timing for specific subgroups.

References
Nederlandse
samenvatting
Acknowledgements
About the author
List of publications
Donders Graduate
School for Cognitive
Neuroscience Series

A

REFERENCES

1. Mullen, E., *Mullen scales of early learning (AGS ed.)*. 1995, Circle Pines, MN: American Guidance Service.
2. Sparrow, S.S., et al., *Vineland adaptive behavior scales (Vineland-II)—2nd edition*. 2005, Mineapolis: Pearson.
3. Bryson, S.E., et al., *The Autism Observation Scale for Infants: scale development and reliability data.* J Autism Dev Disord, 2008. **38**(4): p. 731-8.
4. Gotham, K., A. Pickles, and C. Lord, *Trajectories of autism severity in children using standardized ADOS scores.* Pediatrics, 2012. **130**(5): p. e1278-84.
5. Kanner, L., *Autistic disturbances of affective contact.* Acta Paedopsychiatr, 1968. **35**(4): p. 100-36.
6. Asperger, H., *'Autistic psychopathy' in childhood*, in *Autism and Asperger Syndrome*, U. Frith, Editor. 1991, Cambridge University Press: Cambridge. p. 37-92.
7. Kim, S.H., et al., *Multisite study of new autism diagnostic interview-revised (ADI-R) algorithms for toddlers and young preschoolers.* J Autism Dev Disord, 2013. **43**(7): p. 1527-38.
8. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. 2013, Washington, DC.
9. World Health Organization, *The ICD-10 classification of mental and behavioural disorders : diagnostic criteria for research.* Geneva : World Health Organization, 1993.
10. Boyle, E.A., Y.I. Li, and J.K. Pritchard, *An Expanded View of Complex Traits: From Polygenic to Omnigenic.* Cell, 2017. **169**(7): p. 1177-1186.
11. Chen, J.A., et al., *The emerging picture of autism spectrum disorder: genetics and pathology.* Annu Rev Pathol, 2015. **10**: p. 111-44.
12. de la Torre-Ubieta, L., et al., *Advancing the understanding of autism disease mechanisms through genetics.* Nat Med, 2016. **22**(4): p. 345-61.
13. Richards, C., et al., *Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis.* Lancet Psychiatry, 2015. **2**(10): p. 909-16.
14. Kolesnik, A.M., et al., *Early development of infants with neurofibromatosis type 1: a case series.* Mol Autism, 2017. **8**: p. 62.
15. Tordjman, S., et al., *Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms.* Front Psychiatry, 2014. **5**: p. 53.
16. Christensen, D.L., *Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012 (vol 65, pg 1, 2016).* Mmwr-Morbidity and Mortality Weekly Report, 2016. **65**(15): p. 404-404.
17. Baxter, A.J., et al., *The epidemiology and global burden of autism spectrum disorders.*

Psychol Med, 2015. **45**(3): p. 601-13.

18. Eaves, L.C. and H.H. Ho, *Young adult outcome of autism spectrum disorders.* J Autism Dev Disord, 2008. **38**(4): p. 739-47.
19. Billstedt, E., I.C. Gillberg, and C. Gillberg, *Autism after adolescence: population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood.* J Autism Dev Disord, 2005. **35**(3): p. 351-60.
20. Buescher, A.V., et al., *Costs of autism spectrum disorders in the United Kingdom and the United States.* JAMA Pediatr, 2014. **168**(8): p. 721-8.
21. Fernell, E., M.A. Eriksson, and C. Gillberg, *Early diagnosis of autism and impact on prognosis: a narrative review.* Clin Epidemiol, 2013. **5**: p. 33-43.
22. MacDonald, R., et al., *Assessing progress and outcome of early intensive behavioral intervention for toddlers with autism.* Res Dev Disabil, 2014. **35**(12): p. 3632-44.
23. Dawson, G., *Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder.* Dev Psychopathol, 2008. **20**(3): p. 775-803.
24. Cidav, Z., et al., *Cost Offset Associated With Early Start Denver Model for Children With Autism.* J Am Acad Child Adolesc Psychiatry, 2017. **56**(9): p. 777-783.
25. French, L. and E.M.M. Kennedy, *Annual Research Review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: a systematic review.* J Child Psychol Psychiatry, 2018. **59**(4): p. 444-456.
26. Steiner, A.M., et al., *Practitioner's guide to assessment of autism spectrum disorders in infants and toddlers.* J Autism Dev Disord, 2012. **42**(6): p. 1183-96.
27. Ozonoff, S., et al., *Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study.* Pediatrics, 2011. **128**(3): p. e488-e495.
28. Sandin, S., et al., *The familial risk of autism.* JAMA, 2014. **311**(17): p. 1770-7.
29. Szatmari, P., et al., *Prospective Longitudinal Studies of Infant Siblings of Children With Autism: Lessons Learned and Future Directions.* J Am Acad Child Adolesc Psychiatry, 2016. **55**(3): p. 179-87.
30. Jones, E.J., et al., *Developmental pathways to autism: a review of prospective studies of infants at risk.* Neurosci Biobehav Rev, 2014. **39**: p. 1-33.
31. Zwaigenbaum, L., et al., *Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research.* Pediatrics, 2015. **136 Suppl 1**: p. S10-40.
32. Varcin, K.J. and S.S. Jeste, *The emergence of autism spectrum disorder: insights gained from studies of brain and behaviour in high-risk infants.* Curr Opin Psychiatry, 2017. **30**(2): p. 85-91.
33. Zwaigenbaum, L. and M. Penner, *Autism spectrum disorder: advances in diagnosis and evaluation.* BMJ, 2018. **361**: p. k1674.

34. Yirmiya, N. and T. Charman, *The prodrome of autism: early behavioral and biological signs, regression, peri- and post-natal development and genetics.* J Child Psychol Psychiatry, 2010. **51**(4): p. 432-458.
35. Estes, A., et al., *Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2~years of life.* J Neurodev Disord, 2015. **7**(1): p. 24.
36. Thomas, M.S., et al., *The over-pruning hypothesis of autism.* Dev Sci, 2016. **19**(2): p. 284-305.
37. Jones, W. and A. Klin, *Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism.* Nature, 2013. **504**(7480): p. 427-431.
38. Chawarska, K., S. Macari, and F. Shic, *Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders.* Biol Psychiatry, 2013. **74**(3): p. 195-203.
39. Paul, R., et al., *Out of the mouths of babes: vocal production in infant siblings of children with ASD.* J Child Psychol Psychiatry, 2011. **52**(5): p. 588-98.
40. Flanagan, J.E., et al., *Head lag in infants at risk for autism: a preliminary study.* Am J Occup Ther, 2012. **66**(5): p. 577-85.
41. Di Giorgio, E., et al., *Difference in Visual Social Predispositions Between Newborns at Low- and High-risk for Autism.* Sci Rep, 2016. **6**: p. 26395.
42. Elsabbagh, M. and M.H. Johnson, *Getting answers from babies about autism.* Trends Cogn Sci, 2010. **14**(2): p. 81-7.
43. Gliga, T., et al., *From early markers to neuro-developmental mechanisms of autism.* Dev Rev, 2014. **34**(3): p. 189-207.
44. Rogers, S.J., *What are infant siblings teaching us about autism in infancy?* Autism Res, 2009. **2**(3): p. 125-37.
45. Lord, C., et al., *The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism.* J Autism Dev Disord, 2000. **30**(3): p. 205-23.
46. Zwaigenbaum, L., et al., *Behavioral manifestations of autism in the first year of life.* Int J Dev Neurosci, 2005. **23**(2-3): p. 143-152.
47. Gammer, I., et al., *Behavioural markers for autism in infancy: scores on the Autism Observational Scale for Infants in a prospective study of at-risk siblings.* Infant Behav Dev, 2015. **38**: p. 107-15.
48. Ozonoff, S., et al., *The broader autism phenotype in infancy: when does it emerge?* J Am Acad Child Adolesc Psychiatry, 2014. **53**(4): p. 398-407 e2.
49. Landa, R. and E. Garrett-Mayer, *Development in infants with autism spectrum disorders: a prospective study.* J Child Psychol Psychiatry, 2006. **47**(6): p. 629-638.

50. Chawarska, K., et al., *Parental recognition of developmental problems in toddlers with autism spectrum disorders.* J Autism Dev Disord, 2007. **37**(1): p. 62-72.
51. MacDonald, M., C. Lord, and D. Ulrich, *The relationship of motor skills and adaptive behavior skills in young children with autism spectrum disorders.* Res Autism Spectr Disord, 2013. **7**(11): p. 1383-1390.
52. Leonard, H.C., et al., *Motor development in children at risk of autism: a follow-up study of infant siblings.* Autism, 2014. **18**(3): p. 281-291.
53. Libertus, K., et al., *Limited fine motor and grasping skills in 6-month-old infants at high risk for autism.* Child Dev, 2014. **85**(6): p. 2218-2231.
54. Landa, R., et al., *Developmental trajectories in children with and without autism spectrum disorders: the first 3^years.* Child Dev, 2013. **84**(2): p. 429-442.
55. Bussu, G., et al., *Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis.* J Autism Dev Disord, 2018.
56. Barbaro, J. and C. Dissanayake, *Developmental profiles of infants and toddlers with autism spectrum disorders identified prospectively in a community-based setting.* J Autism Dev Disord, 2012. **42**(9): p. 1939-1948.
57. Toth, K., et al., *Early social, imitation, play, and language abilities of young non-autistic siblings of children with autism.* J Autism Dev Disord, 2007. **37**(1): p. 145-57.
58. Zwaigenbaum, L., et al., *Sex differences in children with autism spectrum disorder identified within a high-risk infant cohort.* J Autism Dev Disord, 2012. **42**(12): p. 2585-96.
59. Salomone, E., et al., *Adaptive Behaviour and Cognitive Skills: Stability and Change from 7 Months to 7 Years in Siblings at High Familial Risk of Autism Spectrum Disorder.* J Autism Dev Disord, 2018.
60. Shiner, R.L., et al., *What is temperament now? Assessing progress in temperament research on the twenty-fifth anniversary of Goldsmith et al. (1987).* Child Development Perspectives, 2012. **.6**(4): p. pp.
61. Nigg, J.T., *Temperament and developmental psychopathology.* J Child Psychol Psychiatry, 2006. **47**(3-4): p. 395-422.
62. Brock, M.E., et al., *Temperament and sensory features of children with autism.* J Autism Dev Disord, 2012. **42**(11): p. 2271-84.
63. Garon, N., et al., *Temperament and its relationship to autistic symptoms in a high-risk infant sib cohort.* J Abnorm Child Psychol, 2009. **37**(1): p. 59-78.
64. Del Rosario, M., et al., *Parent-reported temperament trajectories among infant siblings of children with autism.* J Autism Dev Disord, 2014. **44**(2): p. 381-93.
65. Clifford, S.M., et al., *Temperament in the First 2 Years of Life in Infants at High-Risk for Autism Spectrum Disorders.* Journal of Autism and Developmental Disorders, 2013. **43**(3): p. 673-686.
66. Gomez, C.R. and S. Baird, *Identifying Early Indicators for Autism in Self-Regulation Difficulties.* Focus on Autism and Other Developmental Disabilities, 2005. **.20**(2): p. pp.
67. Bryson, S.E., et al., *A prospective case series of high-risk infants who developed autism.* Journal of Autism and Developmental Disorders, 2007. **37**(1): p. 12-24.
68. Bolton, P.F., et al., *Autism spectrum disorder and autistic traits in the Avon Longitudinal Study of Parents and Children: precursors and early signs.* J Am Acad Child Adolesc Psychiatry, 2012. **51**(3): p. 249-260 e25.
69. Jeste, S.S., J. Frohlich, and S.K. Loo, *Electrophysiological biomarkers of diagnosis and outcome in neurodevelopmental disorders.* Curr Opin Neurol, 2015. **28**(2): p. 110-6.
70. Johnson, M.H., et al., *Annual research review: Infant development, autism, and ADHD--early pathways to emerging disorders.* J Child Psychol Psychiatry, 2015. **56**(3): p. 228-247.
71. de Haan, M., *Infant EEG and event-related potentials*. 2007, Hove, England: Psychology Press.
72. Woodman, G.F., *A brief introduction to the use of event-related potentials in studies of perception and attention.* Atten Percept Psychophys, 2010. **72**(8): p. 2031-46.
73. Walsh, V. and A. Cowey, *Transcranial magnetic stimulation and cognitive neuroscience.* Nat Rev Neurosci, 2000. **1**(1): p. 73-9.
74. Halit, H., M. De Haan, and M. Johnson, *Cortical specialisation for face processing: face-sensitive event-related potential components in 3-and 12-month-old infants.* Neuroimage, 2003. **19**(3): p. 1180-1193.
75. Halit, H., et al., *Face-sensitive cortical processing in early infancy.* Journal of Child Psychology and Psychiatry, 2004. **45**(7): p. 1228-1234.
76. de Haan, M., M.H. Johnson, and H. Halit, *Development of face-sensitive event-related potentials during infancy: a review.* Int J Psychophysiol, 2003. **51**(1): p. 45-58.
77. Johnson, M.H., et al., *The emergence of the social brain network: evidence from typical and atypical development.* Dev Psychopathol, 2005. **17**(3): p. 599-619.
78. Jones, E., G. Dawson, and S. Webb, *Sensory hypersensitivity predicts enhanced attention capture by faces in the early development of ASD.* Developmental cognitive neuroscience, 2018. **29**: p. 11-20.
79. Gliga, T. and G. Dehaene-Lambertz, *Development of a view-invariant representation of the human head.* Cognition, 2007. **102**(2): p. 261-88.
80. Maestro, S., et al., *Attentional Skills During the First 6 Months of Age in Autism Spectrum Disorder.* Journal of the American Academy of Child & Adolescent Psychiatry, 2002.

41(10): p. 1239-1245.

81. Osterling, J. and G. Dawson, *Early recognition of children with autism: A study of first birthday home videotapes.* Journal of Autism and Developmental Disorders, 1994. **24**(3): p. 247-257.
82. Dawson, G., et al., *Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development.* Child Dev, 2002. **73**(3): p. 700-17.
83. Webb, S.J., et al., *ERP evidence of atypical face processing in young children with autism.* J Autism Dev Disord, 2006. **36**(7): p. 881-90.
84. Webb, S.J., et al., *Developmental change in the ERP responses to familiar faces in toddlers with autism spectrum disorders versus typical development.* Child Dev, 2011. **82**(6): p. 1868-86.
85. Grice, S.J., et al., *Neural correlates of eye-gaze detection in young children with autism.* Cortex, 2005. **41**(3): p. 342-353.
86. Jones, E.J., et al., *Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: a longitudinal prospective study of infants at high familial risk.* J Neurodev Disord, 2016. **8**: p. 7.
87. Elsabbagh, M., et al., *Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism.* Curr Biol, 2012. **22**(4): p. 338-342.
88. Dawson, G., et al., *Early behavioral intervention is associated with normalized brain activity in young children with autism.* J Am Acad Child Adolesc Psychiatry, 2012. **51**(11): p. 1150-9.
89. Hahn, T., A.A. Nierenberg, and S. Whitfield-Gabrieli, *Predictive analytics in mental health: applications, guidelines, challenges and perspectives.* Mol Psychiatry, 2017. **22**(1): p. 37-43.
90. Webb, S.J., et al., *The motivation for very early intervention for infants at high risk for autism spectrum disorders.* Int J Speech Lang Pathol, 2014. **16**(1): p. 36-42.
91. Rogers, S.J., et al., *Autism treatment in the first year of life: a pilot study of infant start, a parent-implemented intervention for symptomatic infants.* J Autism Dev Disord, 2014. **44**(12): p. 2981-95.
92. Green, J., et al., *Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial.* Lancet Psychiatry, 2015. **2**(2): p. 133-40.
93. Rosenberg, M.D., B.J. Casey, and A.J. Holmes, *Prediction complements explanation in understanding the developing brain.* Nat Commun, 2018. **9**(1): p. 589.
94. Lai, M.C., et al., *Subgrouping the autism "spectrum": reflections on DSM-5.* PLoS Biol, 2013. **11**(4): p. e1001544.
95. Lord, C., S. Bishop, and D. Anderson, *Developmental trajectories as autism phenotypes.* Am J Med Genet C Semin Med Genet, 2015. **169**(2): p. 198-208.
96. Geurts, H.M., et al., *Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome.* Neuropsychologia, 2008. **46**(13): p. 3030-41.
97. Georgiades, S., S.L. Bishop, and T. Frazier, *Editorial Perspective: Longitudinal research in autism - introducing the concept of 'chronogeneity'.* J Child Psychol Psychiatry, 2017. **58**(5): p. 634-636.
98. Lai, M.C., M.V. Lombardo, and S. Baron-Cohen, *Autism.* Lancet, 2014. **383**(9920): p. 896-910.
99. Vorstman, J.A.S., et al., *Autism genetics: opportunities and challenges for clinical translation.* Nat Rev Genet, 2017. **18**(6): p. 362-376.
100. Jeste, S.S. and D.H. Geschwind, *Disentangling the heterogeneity of autism spectrum disorder through genetic findings.* Nat Rev Neurol, 2014. **10**(2): p. 74-81.
101. Zhao, Y.H. and F.X. Castellanos, *Annual Research Review: Discovery science strategies in studies of the pathophysiology of child and adolescent psychiatric disorders - promises and limitations.* Journal of Child Psychology and Psychiatry, 2016. **57**(3): p. 421-439.
102. Libero, L.E., et al., *Multimodal neuroimaging based classification of autism spectrum disorder using anatomical, neurochemical, and white matter correlates.* Cortex, 2015. **66**: p. 46-59.
103. Lombardo, M.V., et al., *Different functional neural substrates for good and poor language outcome in autism.* Neuron, 2015. **86**(2): p. 567-577.
104. Arbabshirani, M.R., et al., *Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls.* Neuroimage, 2017. **145**(Pt B): p. 137-165.
105. Yahata, N., K. Kasai, and M. Kawato, *Computational neuroscience approach to biomarkers and treatments for mental disorders.* Psychiatry Clin Neurosci, 2017. **71**(4): p. 215-237.
106. Woo, C.W., et al., *Building better biomarkers: brain models in translational neuroimaging.* Nat Neurosci, 2017. **20**(3): p. 365-377.
107. Whelan, R. and H. Garavan, *When optimism hurts: inflated predictions in psychiatric neuroimaging.* Biol Psychiatry, 2014. **75**(9): p. 746-8.
108. Uddin, L.Q., et al., *Salience network-based classification and prediction of symptom severity in children with autism.* JAMA Psychiatry, 2013. **70**(8): p. 869-79.
109. Ingalhalikar, M., et al., *Creating multimodal predictors using missing data: classifying and subtyping autism spectrum disorder.* J Neurosci Methods, 2014. **235**: p. 1-9.
110. Wee, C.Y., et al., *Diagnosis of autism spectrum disorders using regional and interregional morphological features.* Hum Brain Mapp, 2014. **35**(7): p. 3414-30.

111. Emerson, R.W., et al., *Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age.* Sci Transl Med, 2017. **9**(393).
112. Hazlett, H.C., et al., *Early brain development in infants at high risk for autism spectrum disorder.* Nature, 2017. **542**(7641): p. 348-351.
113. Macari, S.L., et al., *Predicting developmental status from 12 to 24 months in infants at risk for Autism Spectrum Disorder: a preliminary report.* J Autism Dev Disord, 2012. **42**(12): p. 2636-47.
114. Chawarska, K., et al., *18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study.* J Am Acad Child Adolesc Psychiatry, 2014. **53**(12): p. 1317-1327.e1.
115. Chen, C.P., et al., *Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism.* Neuroimage Clin, 2015. **8**: p. 238-45.
116. Chen, H., et al., *Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity--A multi-center study.* Prog Neuropsychopharmacol Biol Psychiatry, 2016. **64**: p. 1-9.
117. Iidaka, T., *Resting state functional magnetic resonance imaging and neural network classified autism and control.* Cortex, 2015. **63**: p. 55-67.
118. Quinlan, J.R., *Learning decision tree classifiers.* Acm Computing Surveys, 1996. **28**(1): p. 71-72.
119. Cortes, C. and V. Vapnik, *Support-Vector Networks.* Machine Learning, 1995. **20**(3): p. 273-297.
120. Hinton, G.E. and R.R. Salakhutdinov, *Reducing the dimensionality of data with neural networks.* Science, 2006. **313**(5786): p. 504-7.
121. Insel, T., et al., *Research domain criteria (RDoC): toward a new classification framework for research on mental disorders.* Am J Psychiatry, 2010. **167**(7): p. 748-51.
122. Lombardo, M.V., et al., *Unsupervised data-driven stratification of mentalizing heterogeneity in autism.* Sci Rep, 2016. **6**: p. 35333.
123. Muthen, B. and L.K. Muthen, *Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes.* Alcohol Clin Exp Res, 2000. **24**(6): p. 882-91.
124. Raudenbush, S.W., *Linear mixed models for longitudinal data.* Sociological Methods & Research, 2002. **31**(1): p. 110-118.
125. Muthen, B. and K. Shedden, *Finite mixture modeling with mixture outcomes using the EM algorithm.* Biometrics, 1999. **55**(2): p. 463-9.
126. Groves, A.R., et al., *Linked independent component analysis for multimodal data fusion.* Neuroimage, 2011. **54**(3): p. 2198-217.
127. Smith, S.M., et al., *Advances in functional and structural MR image analysis and implementation as FSL.* Neuroimage, 2004. **23 Suppl 1**: p. S208-19.
128. Jutten, C. and J. Herault, *Blind Separation of Sources .1. An Adaptive Algorithm Based on Neuromimetic Architecture.* Signal Processing, 1991. **24**(1): p. 1-10.
129. Groves, A.R., et al., *Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure.* Neuroimage, 2012. **63**(1): p. 365-80.
130. Douaud, G., et al., *A common brain network links development, aging, and vulnerability to disease.* Proc Natl Acad Sci U S A, 2014. **111**(49): p. 17648-53.
131. Wolfers, T., et al., *Refinement by integration: aggregated effects of multimodal imaging markers on adult ADHD.* J Psychiatry Neurosci, 2017. **42**(6): p. 386-394.
132. Francx, W., et al., *Integrated analysis of gray and white matter alterations in attention-deficit/hyperactivity disorder.* Neuroimage Clin, 2016. **11**: p. 357-67.
133. Green, J., et al., *Randomised trial of a parent-mediated intervention for infants at high risk for autism: longitudinal outcomes to age 3 years.* Journal of Child Psychology and Psychiatry, 2017. **58**(12): p. 1330-1340.
134. Jones, E.J.H., et al., *Developmental pathways to autism: a review of prospective studies of infants at risk.* Neurosci Biobehav Rev, 2014. **39**: p. 1-33.
135. Ozonoff, S., et al., *A prospective study of the emergence of early behavioral signs of autism.* J Am Acad Child Adolesc Psychiatry, 2010. **49**(3): p. 256-66 e1-2.
136. Metz, C.E., *Basic principles of ROC analysis.* Semin Nucl Med, 1978. **8**(4): p. 283-98.
137. Zhou, Y., F. Yu, and T. Duong, *Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning.* PLoS One, 2014. **9**(6): p. e90405.
138. Brian, J., et al., *Clinical assessment of autism in high-risk 18-month-olds.* Autism, 2008. **12**(5): p. 433-56.
139. Judd, C.M., J. Westfall, and D.A. Kenny, *Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem.* J Pers Soc Psychol, 2012. **103**(1): p. 54-69.
140. Bates, D., et al., *Fitting Linear Mixed-Effects Models Using {lme4}.* Journal of Statistical Software, 2015. **67**: p. 1--48.
141. Van Gestel, T., et al., *Bayesian framework for least-squares support vector machine classifiers, gaussian processes, and kernel Fisher discriminant analysis.* Neural Comput, 2002. **14**(5): p. 1115-47.
142. Chawarska, K., et al., *A prospective study of toddlers with ASD: short-term diagnostic*

and cognitive outcomes. J Child Psychol Psychiatry, 2009. **50**(10): p. 1235-45.

143. Wall, D.P., et al., *Use of machine learning to shorten observation-based screening and diagnosis of autism.* Transl Psychiatry, 2012. **2**: p. e100.
144. Perry, A., et al., *Brief report: the Vineland Adaptive Behavior Scales in young children with autism spectrum disorders at different cognitive levels.* J Autism Dev Disord, 2009. **39**(7): p. 1066-78.
145. Liss, M., et al., *Predictors and correlates of adaptive functioning in children with developmental disorders.* J Autism Dev Disord, 2001. **31**(2): p. 219-30.
146. Green, S.A. and A.S. Carter, *Predictors and course of daily living skills development in toddlers with autism spectrum disorders.* J Autism Dev Disord, 2014. **44**(2): p. 256-63.
147. Jasmin, E., et al., *Sensori-motor and daily living skills of preschool children with autism spectrum disorders.* J Autism Dev Disord, 2009. **39**(2): p. 231-41.
148. White, L.K., et al., *Neurobiology and neurochemistry of temperament in children*, in *Handbook of Temperament*, S.R.L. Zentner M, Editor. 2012, 2012: New York, Guilford.
149. Whittle, S., et al., *The neurobiological basis of temperament: towards a better understanding of psychopathology.* Neurosci Biobehav Rev, 2006. **30**(4): p. 511-25.
150. Fox, N.A., *Temperament and early experience form social behavior.* Ann N Y Acad Sci, 2004. **1038**: p. 171-8.
151. Perez-Edgar, K. and N.A. Fox, *Temperament and anxiety disorders.* Child Adolesc Psychiatr Clin N Am, 2005. **14**(4): p. 681-706, viii.
152. Georgiades, S., et al., *A prospective study of autistic-like traits in unaffected siblings of probands with autism spectrum disorder.* Jama Psychiatry, 2013. **70**(1): p. 42-48.
153. Putnam, S.P., L.K. Ellis, and M.K. Rothbart, *The structure of temperament from infancy through adolescence*, in *Advances/proceedings in research on temperament*, A.E.A. Angleitner, Editor. 2001, Pabst Scientist Publisher: Germany. p. 165-182.
154. Gartstein, M.A. and M.K. Rothbart, *Studying infant temperament via the revised Infant Behavior Questionnaire.* Infant Behavior & Development, 2003. **26**(1): p. 64-86.
155. Garon, N., et al., *Temperament and its association with autism symptoms in a high-risk population.* Journal of Abnormal Child Psychology. Aug, 2016(Pagination).
156. Macari, S.L., et al., *Temperamental markers in toddlers with autism spectrum disorder.* J Child Psychol Psychiatry, 2017.
157. Goodman, R., et al., *The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology.* J Child Psychol Psychiatry, 2000. **41**(5): p. 645-55.
158. Rutter, M., A. Bailey, and C. Lord, *SCQ. The Social Communication Questionnaire*. 2003, Los Angeles, CA: Western Psychological Services.

159. Putnam, S.P., M.A. Gartstein, and M.K. Rothbart, *Measurement of fine-grained aspects of toddler temperament: the Early Childhood Behavior Questionnaire.* Infant Behav Dev, 2006. **29**(3): p. 386-401.
160. Lord, C., et al., *Autism Diagnostic Observation Schedule–2nd edition (ADOS-2)*. 2012, Los Angeles, CA: Western Psychological Corporation.
161. Rutter, M., A. Le Couteur, and C. Lord, *ADI-R: Autism diagnostic interview—revised* 2003, Los Angeles, CA: Western Psychological Services.
162. Tabachnik, B.G. and L.S. Fidell, *Using Multivariate Statistics*, ed. t. edition. 2001, Needham Heights, MA: Allyn and Bacon.
163. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate - a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.
164. Tackett, J.L., *Evaluating models of the personality-psychopathology relationship in children and adolescents.* Clin Psychol Rev, 2006. **26**(5): p. 584-99.
165. Picardi, A., et al., *Genetic and environmental influences underlying the relationship between autistic traits and temperament and character dimensions in adulthood* Comprehensive Psychiatry, 2015(Pagination).
166. Johnson, M.H., *Executive function and developmental disorders: the flip side of the coin.* Trends Cogn Sci, 2012. **16**(9): p. 454-457.
167. Loth, E., et al., *The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders.* Mol Autism, 2017. **8**: p. 24.
168. Wan, M.W., et al., *Quality of interaction between at-risk infants and caregiver at 12-15 months is associated with 3-year autism outcome.* J Child Psychol Psychiatry, 2013. **54**(7): p. 763-71.
169. Gagne, J.R., et al., *Deriving childhood temperament measures from emotion-eliciting behavioral episodes: scale construction and initial validation.* Psychol Assess, 2011. **23**(2): p. 337-53.
170. Rothbart, M.K., *Longitudinal Observation of Infant Temperament.* Developmental Psychology, 1986. **22**(3): p. 356-365.
171. Voelker, P., et al., *Variations in catechol-O-methyltransferase gene interact with parenting to influence attention in early development.* Neuroscience, 2009. **164**(1): p. 121-30.
172. Sheese, B.E., et al., *Parenting quality interacts with genetic variation in dopamine receptor D4 to influence temperament in early childhood.* Dev Psychopathol, 2007. **19**(4): p. 1039-46.
173. Barbaro, J. and C. Dissanayake, *Autism spectrum disorders in infancy and toddlerhood:*

a review of the evidence on early signs, early identification tools, and early diagnosis. Journal of Developmental & Behavioral Pediatrics, 2009. **30**(5): p. 447-459.

174. Jones, E., et al., *Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: a longitudinal prospective study of infants at high familial risk.* Journal of neurodevelopmental disorders, 2016. **8**(1): p. 7.
175. Elsabbagh, M. and M.H. Johnson, *Autism and the social brain: the first-year puzzle.* Biological psychiatry, 2016. **80**(2): p. 94-99.
176. Snaedal, J., et al., *Diagnostic accuracy of statistical pattern recognition of electroencephalogram registration in evaluation of cognitive impairment and dementia.* Dementia and geriatric cognitive disorders, 2012. **34**(1): p. 51-60.
177. Bedford, R., et al., *Neurocognitive and observational markers: prediction of autism spectrum disorder from infancy to mid-childhood.* Mol Autism, 2017. **8**: p. 49.
178. Kumar, R. and A. Indrayan, *Receiver operating characteristic (ROC) curve for medical researchers.* Indian pediatrics, 2011. **48**(4): p. 277-287.
179. Green, J., et al., *Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial.* The Lancet Psychiatry, 2015. **2**(2): p. 133-140.
180. Eldridge, J., et al., *Robust features for the automatic identification of autism spectrum disorder in children.* J Neurodev Disord, 2014. **6**(1): p. 12.
181. J.H., J.E., et al., *Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention.* Autism Research, 2017. **10**(5): p. 961-972.
182. Jones, E.J., et al., *Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention.* Autism Research, 2017. **10**(5): p. 961-972.
183. Johannesson, G.H., et al., *Combined electronic structure and evolutionary search approach to materials design.* Phys Rev Lett, 2002. **88**(25 Pt 1): p. 255506.
184. Dash, M.a.L., H., *Feature selection for classification.* Intelligent Data Analysis, 1997. **1**: p. 131-156.
185. Back, T., *Evolution strategies: An alternative evolutionary algorithm.* Artificial Evolution, 1996. **1063**: p. 3-20.
186. Chang, C.-C.a.L., Chih-Jen, *LIBSVM: A library for support vector machines.* ACM Transactions on Intelligent Systems and Technology, 2011. **2**(3): p. 27:1--27: 27.
187. Golland, P. and B. Fischl, *Permutation tests for classification: towards statistical significance in image-based studies.* Inf Process Med Imaging, 2003. **18**: p. 330-41.
188. Howlin, P. and I. Magiati, *Autism spectrum disorder: outcomes in adulthood.* Curr Opin Psychiatry, 2017. **30**(2): p. 69-76.
189. Howlin, P., et al., *Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children.* J Am Acad Child Adolesc Psychiatry, 2013. **52**(6): p. 572-81 e1.
190. Howlin, P., et al., *Cognitive and language skills in adults with autism: a 40-year follow-up.* J Child Psychol Psychiatry, 2014. **55**(1): p. 49-58.
191. Szatmari, P., et al., *Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder.* JAMA Psychiatry, 2015. **72**(3): p. 276-83.
192. Farmer, C., et al., *Classifying and characterizing the development of adaptive behavior in a naturalistic longitudinal study of young children with autism.* J Neurodev Disord, 2018. **10**(1): p. 1.
193. Flanagan, H.E., et al., *Stability and Change in the Cognitive and Adaptive Behaviour Scores of Preschoolers with Autism Spectrum Disorder.* J Autism Dev Disord, 2015. **45**(9): p. 2691-703.
194. Bal, V.H., et al., *Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age.* Autism, 2015. **19**(7): p. 774-84.
195. Baghdadli, A., et al., *Developmental trajectories of adaptive behaviors from early childhood to adolescence in a cohort of 152 children with autism spectrum disorders.* J Autism Dev Disord, 2012. **42**(7): p. 1314-25.
196. Sparrow S, B.D. and C. D., *Vineland adaptive behavior sscale: second edition.* Shoreview, MN: American Guidance Service, 2005.
197. Zwaigenbaum, L., et al., *Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research.* Pediatrics, 2015. **136 Suppl 1**: p. S60-81.
198. Sacrey, L.R., et al., *Developmental trajectories of adaptive behavior in autism spectrum disorder: a high-risk sibling cohort.* J Child Psychol Psychiatry, 2018.
199. Bolte, S. and F. Poustka, *The relation between general cognitive level and adaptive behavior domains in individuals with autism with and without co-morbid mental retardation.* Child Psychiatry Hum Dev, 2002. **33**(2): p. 165-72.
200. Kanne, S.M., et al., *The role of adaptive behavior in autism spectrum disorders: implications for functional outcome.* J Autism Dev Disord, 2011. **41**(8): p. 1007-18.
201. Klin, A., et al., *Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the Vineland and the ADOS.* J Autism Dev Disord, 2007. **37**(4): p. 748-59.
202. Gammer, I., et al., *Behavioural markers for autism in infancy: scores on the Autism*

Observational Scale for Infants in a prospective study of at-risk siblings. Infant Behav Dev, 2015. **38**: p. 107-115.

203. Mullen, E.M., *Mullen Scale of Early Learning: AGS edition.* Circle Pines, MN: American Guidance Service Publishing, 1995.
204. Lord, C.R., M.; DiLavore, P.; Risi, S.; Gotham, K.; Bishop, S., *Autism diagnostic observation schedule–2nd edition (ADOS-2).* 2012, Los Angeles, CA: Western Psychological Corporation.
205. Rutter, M., Bailey, A., Lord, C., *The social communication questionnaire: manual.* 2003, Western Psychological Services.
206. IBM Corp. *IBM SPSS software*. 02-11-2018]; Available from: https://www.ibm.com/analytics/spss-statistics-software.
207. Nagin, D.S. and C.L. Odgers, *Group-Based Trajectory Modeling in Clinical Research.* Annual Review of Clinical Psychology, Vol 6, 2010. **6**: p. 109-138.
208. Proust-Lima, C., V. Philipps, and B. Liquet, *Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm.* Journal of Statistical Software, 2017. **78**(2): p. 1-56.
209. Geschwind, D.H. and P. Levitt, *Autism spectrum disorders: developmental disconnection syndromes.* Curr Opin Neurobiol, 2007. **17**(1): p. 103-11.
210. Wolff, J.J., et al., *Differences in White Matter Fiber Tract Development Present From 6 to 24 Months in Infants With Autism.* American Journal of Psychiatry, 2012. **169**(6): p. 589-600.
211. Solso, S., et al., *Diffusion Tensor Imaging Provides Evidence of Possible Axonal Overconnectivity in Frontal Lobes in Autism Spectrum Disorder Toddlers.* Biological Psychiatry, 2016. **79**(8): p. 676-684.
212. Piven, J., J.T. Elison, and M.J. Zylka, *Toward a conceptual framework for early brain and behavior development in autism.* Mol Psychiatry, 2017. **22**(10): p. 1385-1394.
213. Meredith, R.M., *Sensitive and critical periods during neurotypical and aberrant neurodevelopment: a framework for neurodevelopmental disorders.* Neurosci Biobehav Rev, 2015. **50**: p. 180-8.
214. Lai, M.C., et al., *Sex/gender differences and autism: setting the scene for future research.* J Am Acad Child Adolesc Psychiatry, 2015. **54**(1): p. 11-24.
215. Mandic-Maravic, V., et al., *Sex differences in autism spectrum disorders: does sex moderate the pathway from clinical symptoms to adaptive behavior?* Sci Rep, 2015. **5**: p. 10418.
216. Charman, T., et al., *IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP).* Psychol Med, 2011. **41**(3): p. 619-27.

217. Bolck, A., M. Croon, and J. Hagenaars, *Estimating latent structure models with categorical variables: One-step versus three-step estimators.* Political Analysis, 2004. **12**(1): p. 3-27.
218. Jones, W. and A. Klin, *Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism.* Nature, 2013. **504**(7480): p. 427-31.
219. Shen, M.D. and J. Piven, *Brain and behavior development in autism from birth through infancy.* Dialogues Clin Neurosci, 2017. **19**(4): p. 325-333.
220. Elsabbagh, M., et al., *Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood.* Biol Psychiatry, 2013. **74**(3): p. 189-94.
221. Shic, F., S. Macari, and K. Chawarska, *Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder.* Biol Psychiatry, 2014. **75**(3): p. 231-7.
222. Elison, J.T., et al., *White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism.* Am J Psychiatry, 2013. **170**(8): p. 899-908.
223. Shen, M.D., et al., *Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder.* Brain, 2013. **136**(Pt 9): p. 2825-35.
224. Shen, M.D., et al., *Increased Extra-axial Cerebrospinal Fluid in High-Risk Infants Who Later Develop Autism.* Biol Psychiatry, 2017. **82**(3): p. 186-193.
225. Hendry, A., et al., *Developmental change in look durations predicts later effortful control in toddlers at familial risk for ASD.* J Neurodev Disord, 2018. **10**(1): p. 3.
226. Mundy, P. and L. Newell, *Attention, Joint Attention, and Social Cognition.* Curr Dir Psychol Sci, 2007. **16**(5): p. 269-274.
227. Charman, T., et al., *Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): A baby siblings research consortium (BSRC) study.* Autism Res, 2017. **10**(1): p. 169-178.
228. Barger, B.D., J.M. Campbell, and J.D. McDonough, *Prevalence and onset of regression within autism spectrum disorders: a meta-analytic review.* J Autism Dev Disord, 2013. **43**(4): p. 817-28.
229. Baird, G., et al., *Regression, Developmental Trajectory and Associated Problems in Disorders in the Autism Spectrum: The SNAP Study.* Journal of Autism and Developmental Disorders, 2008. **38**(10): p. 1827-1836.
230. Ozonoff, S., et al., *Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study.* J Child Psychol Psychiatry, 2015. **56**(9): p. 988-98.
231. Bacon, E.C., et al., *Rethinking the idea of late autism spectrum disorder onset.* Dev Psychopathol, 2017: p. 1-17.
232. Vivanti, G., C. Dissanayake, and A.T. Victorian, *Outcome for Children Receiving the Early Start Denver Model Before and After 48 Months.* J Autism Dev Disord, 2016. **46**(7): p.

2441-9.

233. Jones, E.J.H., et al., *Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention.* Autism Res, 2017. **10**(5): p. 961-972.
234. Sacrey, L.A., et al., *Impairments to visual disengagement in autism spectrum disorder: a review of experimental studies from infancy to adulthood.* Neurosci Biobehav Rev, 2014. **47**: p. 559-77.
235. Ozonoff, S., et al., *How early do parent concerns predict later autism diagnosis?* J Dev Behav Pediatr, 2009. **30**(5): p. 367-75.
236. Chevallier, C., et al., *The social motivation theory of autism.* Trends Cogn Sci, 2012. **16**(4): p. 231-9.
237. Zwaigenbaum, L., et al., *Stability of diagnostic assessment for autism spectrum disorder between 18 and 36 months in a high-risk cohort.* Autism Res, 2016. **9**(7): p. 790-800.
238. Brian, J., et al., *Stability and change in autism spectrum disorder diagnosis from age 3 to middle childhood in a high-risk sibling cohort.* Autism, 2016. **20**(7): p. 888-92.
239. Lundstrom, S., et al., *Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation.* Arch Gen Psychiatry, 2012. **69**(1): p. 46-52.
240. Robinson, E.B., et al., *Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%).* Arch Gen Psychiatry, 2011. **68**(11): p. 1113-21.
241. Uher, R. and M. Rutter, *Basing psychiatric classification on scientific foundation: problems and prospects.* Int Rev Psychiatry, 2012. **24**(6): p. 591-605.
242. Ecker, C., et al., *Association Between the Probability of Autism Spectrum Disorder and Normative Sex-Related Phenotypic Diversity in Brain Structure.* Jama Psychiatry, 2017. **74**(4): p. 329-338.
243. Marquand, A.F., et al., *Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies.* Biological Psychiatry, 2016. **80**(7): p. 552-561.
244. Campbell, D.J., et al., *Gaze response to dyadic bids at 2 years related to outcomes at 3 years in autism spectrum disorders: a subtyping analysis.* J Autism Dev Disord, 2014. **44**(2): p. 431-42.
245. Hagenaars, J.A., *Categorical causal modeling - Latent class analysis and directed log-linear models with latent variables.* Sociological Methods & Research, 1998. **26**(4): p. 436-486.
246. Munson, J., et al., *Evidence for latent classes of IQ in young children with autism spectrum disorder.* Am J Ment Retard, 2008. **113**(6): p. 439-52.
247. Bishop-Fitzpatrick, L., et al., *Characterizing Objective Quality of Life and Normative Outcomes in Adults with Autism Spectrum Disorder: An Exploratory Latent Class Analysis.* J Autism Dev Disord, 2016. **46**(8): p. 2707-19.
248. James, R.J., et al., *The Latent Structure of Autistic Traits: A Taxometric, Latent Class and Latent Profile Analysis of the Adult Autism Spectrum Quotient.* J Autism Dev Disord, 2016. **46**(12): p. 3712-3728.
249. Wiggins, L.D., et al., *Homogeneous Subgroups of Young Children with Autism Improve Phenotypic Characterization in the Study to Explore Early Development.* J Autism Dev Disord, 2017. **47**(11): p. 3634-3645.
250. Hagenaars, J.A.M., A.L., *Applied Latent Class Analysis*, ed. C.U. Press. 2009.
251. Visser, J.C., et al., *Variation in the Early Trajectories of Autism Symptoms Is Related to the Development of Language, Cognition, and Behavior Problems.* J Am Acad Child Adolesc Psychiatry, 2017. **56**(8): p. 659-668.
252. Venker, C.E., et al., *Trajectories of autism severity in early childhood.* J Autism Dev Disord, 2014. **44**(3): p. 546-63.
253. Fountain, C., A.S. Winter, and P.S. Bearman, *Six Developmental Trajectories Characterize Children With Autism.* Pediatrics, 2012. **129**(5): p. E1112-E1120.
254. Messinger, D., et al., *Beyond autism: a baby siblings research consortium study of high-risk children at three years of age.* J Am Acad Child Adolesc Psychiatry, 2013. **52**(3): p. 300-308 e1.
255. Vismara, L.A. and S.J. Rogers, *The Early Start Denver Model A Case Study of an Innovative Practice.* Journal of Early Intervention, 2008. **31**(1): p. 91-108.
256. Dawson, G., et al., *Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model.* Pediatrics, 2010. **125**(1): p. e17-23.
257. Eapen, V. and R.A. Clarke, *Autism spectrum disorders: from genotypes to phenotypes.* Front Hum Neurosci, 2014. **8**: p. 914.
258. Lenroot, R.K. and P.K. Yeung, *Heterogeneity within Autism Spectrum Disorders: What have We Learned from Neuroimaging Studies?* Front Hum Neurosci, 2013. **7**: p. 733.
259. Georgiades, S., et al., *Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach.* J Child Psychol Psychiatry, 2013. **54**(2): p. 206-15.
260. Broadstock, M., S. Michie, and T. Marteau, *Psychological consequences of predictive genetic testing: a systematic review.* Eur J Hum Genet, 2000. **8**(10): p. 731-8.
261. Bailey, D.B., Jr., et al., *Maternal Consequences of the Detection of Fragile X Carriers in Newborn Screening.* Pediatrics, 2015. **136**(2): p. e433-40.
262. Chung, J., et al., *Twenty-year follow-up of newborn screening for patients with muscular dystrophy.* Muscle Nerve, 2016. **53**(4): p. 570-8.

263. Russell, G. and B. Norwich, *Dilemmas, diagnosis and de-stigmatization: parental perspectives on the diagnosis of autism spectrum disorders.* Clin Child Psychol Psychiatry, 2012. **17**(2): p. 229-45.
264. Bailey, D.B., Jr., et al., *Ethical, legal, and social concerns about expanded newborn screening: fragile X syndrome as a prototype for emerging issues.* Pediatrics, 2008. **121**(3): p. e693-704.
265. Walsh, P., et al., *In search of biomarkers for autism: scientific, social and ethical challenges.* Nat Rev Neurosci, 2011. **12**(10): p. 603-12.
266. Cawthorpe, D., *Comprehensive Description of Comorbidity for Autism Spectrum Disorder in a General Population.* Perm J, 2017. **21**.
267. Matson, J.L. and R.L. Goldin, *Comorbidity and autism: Trends, topics and future directions.* Research in Autism Spectrum Disorders, 2013. **7**(10): p. 1228-1233.
268. Matson, J.L. and P.E. Cervantes, *Commonly studied comorbid psychopathologies among persons with autism spectrum disorder.* Res Dev Disabil, 2014. **35**(5): p. 952-62.
269. Simonoff, E., et al., *Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample.* J Am Acad Child Adolesc Psychiatry, 2008. **47**(8): p. 921-9.
270. Gargaro, B.A., et al., *Autism and ADHD: how far have we come in the comorbidity debate?* Neurosci Biobehav Rev, 2011. **35**(5): p. 1081-8.
271. Mannion, A. and G. Leader, *An investigation of comorbid psychological disorders, sleep problems, gastrointestinal symptoms and epilepsy in children and adolescents with autism spectrum disorder: A two year follow-up.* Research in Autism Spectrum Disorders, 2016. **22**: p. 20-33.
272. Matson, J.L. and L.W. Williams, *Differential diagnosis and comorbidity: distinguishing autism from other mental health issues.* Neuropsychiatry, 2013. **3**(2): p. 233-243.
273. Leyfer, O.T., et al., *Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders.* Journal of Autism and Developmental Disorders, 2006. **36**(7): p. 849-861.
274. Horvath, K. and J.A. Perman, *Autistic disorder and gastrointestinal disease.* Curr Opin Pediatr, 2002. **14**(5): p. 583-7.
275. Tuchman, R. and I. Rapin, *Epilepsy in autism.* Lancet Neurol, 2002. **1**(6): p. 352-8.
276. Sokolova, E., et al., *A Causal and Mediation Analysis of the Comorbidity Between Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD).* J Autism Dev Disord, 2017. **47**(6): p. 1595-1604.
277. Chiang, H.L. and S.S. Gau, *Comorbid psychiatric conditions as mediators to predict later social adjustment in youths with autism spectrum disorder.* J Child Psychol Psychiatry, 2016. **57**(1): p. 103-111.
278. Doshi-Velez, F., Y. Ge, and I. Kohane, *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis.* Pediatrics, 2014. **133**(1): p. e54-63.
279. Zhang, M.L. and Z.H. Zhou, *A Review on Multi-Label Learning Algorithms.* Ieee Transactions on Knowledge and Data Engineering, 2014. **26**(8): p. 1819-1837.
280. Pan, S.J. and Q.A. Yang, *A Survey on Transfer Learning.* Ieee Transactions on Knowledge and Data Engineering, 2010. **22**(10): p. 1345-1359.
281. Wolfers, T., et al., *From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics.* Neurosci Biobehav Rev, 2015. **57**: p. 328-49.
282. Jeste, S.S., et al., *Symptom profiles of autism spectrum disorder in tuberous sclerosis complex.* Neurology, 2016. **87**(8): p. 766-72.
283. Fernandez, B.A. and S.W. Scherer, *Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach.* Dialogues Clin Neurosci, 2017. **19**(4): p. 353-371.
284. Kong, A., et al., *The nature of nurture: Effects of parental genotypes.* Science, 2018. **359**(6374): p. 424-428.
285. Gardener, H., D. Spiegelman, and S.L. Buka, *Prenatal risk factors for autism: comprehensive meta-analysis.* British Journal of Psychiatry, 2009. **195**(1): p. 7-14.
286. Durkin, M.S., et al., *Advanced parental age and the risk of autism spectrum disorder.* Am J Epidemiol, 2008. **168**(11): p. 1268-76.
287. Field, S.S., *Interaction of genes and nutritional factors in the etiology of autism and attention deficit/hyperactivity disorders: a case control study.* Med Hypotheses, 2014. **82**(6): p. 654-61.
288. Agrawal, S., et al., *Prevalence of Autism Spectrum Disorder in Preterm Infants: A Meta-analysis.* Pediatrics, 2018. **142**(3).

NEDERLANDSE SAMENVATTING

Autisme Spectrum Stoornis (ASS) is een meestal levenslange conditie die ontstaat in de eerste jaren na de geboorte. Een klinische diagnose wordt echter gewoonlijk pas op latere leeftijd gesteld, meestal tussen 4 en 5 jaar [26]. Het is de huidige gewoonte dat gedragsinterventies alleen worden uitgevoerd nadat een klinische diagnose van ASS is vastgesteld. Deze gedragsinterventies zorgen voor een betere ontwikkeling van sociale en communicatieve vaardigheden en voor een vermindering van onaangepast gedrag. Het uiteindelijke doel is dat het latere dagelijks functioneren verbetert [21, 22]. Echter, eerder onderzoek suggereert dat een vroege interventie effectiever is dan een latere behandeling [91, 92, 232, 233], vooral wanneer die wordt geïnitieerd in het eerste levensjaar waarin de kernsymptomen van ASS nog niet zijn opgetreden [90]. Om vroege interventies gericht op de persoon zelf mogelijk te maken, is het belangrijk om de kinderen die vroeg in hun leven hulp nodig hebben te identificeren. Dit proefschrift behandelt twee hoofdthema's in het onderzoek naar de ASS-ontwikkeling in de kindertijd: het voorspellen van de ASS diagnose op latere leeftijd bij het individuele kind, en het inzicht in de heterogeniteit van ASS door te onderzoeken welke cognitieve en neurale mechanismen aan de ASS-ontwikkeling ten grondslag liggen. Het gemeenschappelijke punt was de overgang van bevindingen bij groepen van kinderen naar bevindingen bij het individuele kind.

VOORSPELLEN VAN DE ASS UITKOMST

Op basis van onderzoek naar groepsverschillen tussen kinderen met of zonder ASS kunnen we weinig over het individuele kind vertellen want de overlap in individuele variatie tussen groepen kan aanzienlijk zijn. Dus, het eerste deel van dit proefschrift (hoofdstuk 2 tot 4) concentreert zich op het vertalen van groepsverschillen naar het voorspellen van de ASS uitkomst op jonge leeftijd bij het individuele kind. Deze hoofdstukken gaan vooral over het onderzoek van vroege kenmerken en voorlopers van ASS in de eerste twee levensjaren. Naast vroege identificeren van ASS, biedt het voorspellen van de individuele ASS uitkomst (dus of op de leeftijd van drie jaar sprake is van ASS danwel van een "normale" ontwikkeling) ook meer inzicht in de onderliggende cognitieve en neurale mechanismen in de ontwikkeling van ASS door de identificatie van potentiële biomarkers voor de stoornis. Eerdere studies hebben al geprobeerd de individuele ASS uitkomst in de eerste 2 levensjaren te voorspellen. Aangezien ASS gekenmerkt wordt door een sterke klinische, biologische en etiologische heterogeniteit [94], is echter een vroege detectie bij het individuele kind moeilijk en onzeker. Om deze beperking te ondervangen, combineerde ik verschillende soorten data in multivariate modellen voor het voorspellen van latere klinische uitkomsten door gebruik te maken van machine-learning technieken voor gesuperviseerd leren.

In Hoofdstuk 2 gebruikte ik een Support Vector Machine classificeerder (SVM) om de individuele ASD-uitkomst bij kinderen met een hoog risico van een ASS diagnose, als broers en zussen van kinderen met een ASS diagnose (HR kinderen), te voorspellen met behulp van metingen van het ontwikkelingsniveau, het adaptief functioneren en vroege [78] ASS-symptomen bij 8 en 14 maanden. Verder heb ik groepstrajecten van het ontwikkelingsniveau en het adaptief functioneren tussen 8 en 36 maanden bij verschillende klinische uitkomstgroepen gekarakteriseerd: kinderen met een laag risico op een ASS diagnose (LR); HR kinderen met een uitkomst van typische ontwikkeling (HR-Typisch), atypische ontwikkeling (HR-Atypisch) of ASS (HR-ASS) bij 36 maanden. LR en HR-typisch groepen hadden een hoger ontwikkelingsniveau en een hoger dagelijkse functioneren dan HR-Atypisch en HR-ASS in de loop der tijd. Bij het individuele HR kind was het voorspellen van de ASS uitkomst mogelijk met een Area Under the Curve (AUC) van 71% met behulp van metingen van alledaagse vaardigheden op 14 levensmaanden. De classificeerder had een veel lagere positieve voorspellende waarde (PVW = 29%) dan de negatieve voorspellende waarde (NVW = 92%). Dit betekent dat informatie over alledaagse vaardigheden tot een betere voorspelling leidt van kinderen die geen ASS diagnose krijgen. Dit kan voor de klinische praktijk handig zijn door de identificatie van kinderen die niet intensief behoeven te worden geobserveerd.

In Hoofdstuk 3 onderzocht ik het temperament als een vroeg kenmerk van ASS door onderzoek van de door de ouders gerapporteerde vragenlijsten bij HR en LR kinderen bij 8, 14 en 24 maanden. Net als bij hoofdstuk 2 heb ik ontwikkelingstrajecten van de verschillende groepen gekarakteriseerd en heb ik een SVM classificeerder gebruikt om de individuele ASS-uitkomst op 36 levensmaanden bij HR kinderen te voorspellen. Onze bevindingen tonen verschillen tussen klinische uitkomstgroepen, met meer atypisch temperament bij HR-ASS kinderen, respectievelijk gevolgd door HR-Atypische, HR-Typische en LR kinderen. Verder was het mogelijk om de ASD-uitkomst bij het individuele kind te voorspellen met een AUC van 71% met behulp van maten van zelfregulatie op 24 levensmaanden. Hier was ook de positief voorspellende waarde laag (PVW = 29%), dus HR-ASS kinderen konden niet accuraat worden geïdentificeerd. Omgekeerd gaf een negatief voorspellende waarde van 96% aan dat HR kinderen zonder ASS wel voldoende accuraat konden worden voorspeld. Dit suggereert dat temperament niet bijdraagt aan de vroege identificatie van kinderen die een ASS diagnose op latere leeftijd krijgen, maar wel bijdraagt aan de herkenning van het individuele kind dat geen ASS diagnose krijgt.

In Hoofdstuk 4 onderzocht ik bij 8 maanden groeps- en individuele associaties van de neurale gevoeligheid voor visuele ruis (een “gescrambeld” gezicht) en gezichten met statisch en dynamisch oogcontact bij HR kinderen met een ASS uitkomst op de leeftijd van

36 maanden. Het voorspellen van de individuele ASS uitkomst was mogelijk met ongeveer 80% AUC, 81% PVW en 71% NVW, door een SVM classificeerder die gebruik maakte van de snelheid en de amplitude van de neurale respons op vroege en latere stadia van de gezichtsverwerking. Dit patroon kwam overeen met de resultaten over klinische uitkomstgroepen en geeft aan dat kinderen die ASS ontwikkelen problemen ondervinden als gevolg van afwijkingen in de neurale verwerking van gezichten die al op de kindertijd ontstaan.

HETEROGENITEIT VAN ASS

Een andere beperking van het huidige onderzoek naar ASS op de kindertijd is dat het berust op het vergelijken van cases en controles. Het case-control design is namelijk gebaseerd op klinische labels die zijn gedefinieerd op het moment van de diagnose. Deze labels houden geen rekening met de heterogeniteit van ASS [121] en hebben misschien zelfs weinig tot geen betekenis in de vroege ontwikkeling. Dus, in het tweede deel van dit proefschrift (hoofdstuk 5 en 6) gebruikte ik een niet-gestuurde vorm van machine learning om onbekende structuren in data te ontdekken en de relatie met de huidige categorische uitkomsten te onderzoeken. Deze onbekende structuren in data kunnen een beter en completer inzicht bieden in de diversiteit en heterogeniteit van ASS door de identificatie van subgroepen bij kinderen met ASS of onderliggende cognitieve en hersenmechanismen.

In Hoofdstuk 5 heb ik verschillende groepen bij HR en LR kinderen geïdentificeerd op basis van vroege ontwikkeling van adaptief functioneren tussen 8 en 36 levensmaanden door gebruik te maken van parallelle Growth Mixture Modelling (GMM). Ik observeerde drie verschillende groeitrajecten: aanvankelijk hoge maar vervolgens dalende scores over alle Vineland schalen (8,3%); stabiele trajecten rond leeftijdsgebonden normen (73,8%); en aanvankelijk lage maar vervolgens toenemend trajecten die een stabiel gemiddeld score op alle schalen op de leeftijd van twee jaar bereiken (17,9%). Het dalende traject van het adaptief functioneren had een significant verhoogd risico [OR = 4.40 (CI: 1.90; 12.98)] voor een ASS uitkomst in vergelijking met de andere trajecten. Kinderen in deze groep hadden ook meer symptomen in het sociale, communicatie- en repetitieve gedragsdomein op 36 levensmaanden. Verder observeerde ik de opkomst van een patroon dat een discrepantie in de ontwikkeling tussen het adaptief functioneren en het cognitief niveau toont. In het bijzonder vertoonde de verbeterende klasse in het adaptief functioneren stabiele trajecten rond gemiddelde scores van cognitieve ontwikkeling.

Ten slotte ontdekte ik in hoofdstuk 6 processen die van invloed op meerdere ontwikkelingsdomeinen zijn door gebruik te maken van niet-gestuurde machine learning.

Daarbij selecteerde ik de processen die significant geassocieerd zijn met de klinische classificatie bij 36 maanden. In het bijzonder analyseerde ik het ontwikkelingsniveau, het adaptief functioneren en vroege symptomen van ASS tussen 8 en 36 levensmaanden, en de neurale gevoeligheid voor kijkrichtingen bij 8 maanden bij HR en LR kinderen. Ik identificeerde twee onafhankelijke, longitudinale processen en een proces op jonge levenstijd (8 levensmaanden) die coherente effecten over meerdere ontwikkelingsdomeinen lieten zien. Het vroege proces was met niet-ASS uitkomsten geassocieerd en het werd gekarakteriseerd door een hoog adaptief en cognitief functioneren en een laag niveau van ASS symptomen in verband met verminderde aandacht, maar grotere betrokkenheid bij blikverschuivingen en verminderde betrokkenheid bij niet-sociale stimuli. Vervolgens werd één longitudinaal proces geassocieerd met niet-ASD uitkomsten en werd het gekarakteriseerd door een toenemend cognitief en adaptief functioneren en lage niveaus van ASS symptomen. Het andere longitudinaal proces was geassocieerd met een ASS uitkomst en het wees een stagnatie in cognitief functioneren na 24 levensmaanden aan.

CONCLUSIE

Dit proefschrift wijst een diffuus patroon van neurale responsen op gezichten en visuele ruis op 8 levensmaanden aan als een mogelijke voorloper en voorspeller van de ASS uitkomst op de leeftijd van 36 maanden. Daarbij is er sprake van een goede voorspellende nauwkeurigheid in het eerste levensjaar, wanneer gedragssignalen en symptomen van ASS nog steeds niet gevoelig en specifiek genoeg zijn voor het voorspellen van een ASS diagnose bij het individuele kind. Dit wijst op een algemene verandering in de neurale verwerking van gezichten als een indicator voor latere ASS. Bovendien hebben geavanceerde statistische analyses ons in staat gesteld om de kracht van het prospectieve onderzoeksdesign volledig te benutten om subgroepen bij HR en LR kinderen te identificeren, en onderliggende processen te ontdekken die vroeg in de ontwikkeling samenwerken en geassocieerd zijn met een ASS uitkomst. Het toekomstig onderzoek dient voorgezet te worden om homogene subgroepen bij kinderen met ASS te identificeren en biomarkers voor die subgroepen vast te stellen en te valideren. Het doel is om voorspellingen op basis van responsen op gezichten te combineren met bestaande klinische expertise. Dit zal leiden tot een eerdere en betrouwbaardere detectie van ASS en het mogelijk maken om de beste behandeling voor specifieke subgroepen te bepalen. en ook vast te stellen wat de beste leeftijd is om die behandeling te starten.

ACKNOWLEDGEMENTS

Finally, the part that everyone is waiting for, the one that is supposed to be the easiest after all this hard work, but is actually the most difficult for me to write. I must admit the Dutch summary put me seriously to the test, though. Although I'm really grateful to all the people who helped me throughout the hard path leading to this point, it's difficult for me to go beyond the simple "thank you" as that would mean showing off my feelings. But I'll try my best here.

The first person I'd like to thank is Jan. With no doubts, you taught me a lot as a researcher and as a person. Coming from a different background, you introduced me to the autism research and helped me overcoming all the challenges this new field of research put me through. I learned from you how to think differently, and how patience and diplomacy are important to reach our final goals. I respect and admire the way you manage so many projects, supervising so many students but still finding the time and the right way to make each one of us young researchers enthusiastic about our work. I find really inspiring to see how, despite your great knowledge in neuroscience and autism, you are still so curious and eager to learn more. You always inspired me with new ideas, but also encouraged me to find my own way. Bedankt voor alles!

Then I want to thank Christian. Although we have not worked together as much as I would have loved to in the past three years, I've learned from our meetings how to properly look at the data and do something useful with it, and to be always critical. I admire the way you make all look so simple and natural, explaining the core concepts so clearly and always with a funny note. I'm looking forward to working more together for the new project as I'm sure I'll learn a lot! In particular, I loved working on our FLICA paper, learning that new technique and thinking about all the possible applications and implications for our research. And for this, special thanks also go to Alberto. We have had many meetings and you always found the time to explain me the main concepts of our analyses. Most importantly, you did that with a big smile. Muchas gracias!

Besides my promotors from Nijmegen, I want to thank all my colleagues and people from the Donders Institute for their daily contribution at creating such a vibrant and stimulating working environment. I learned a lot from all the meetings and presentations. In particular, I want to thank Mirjam for collaborating with me on the temperament study. It's been nice to exchange views about different analytic approaches, with the same aim to predict autism at an early age.

A great thank you goes to Emily for co-supervising me. You welcomed me in London the first year of my PhD and introduced me to the BASIS dataset. Probably due to my background, it's unusual to me to see a young woman so successful in research and I admire you for that. Especially now that I'm newly mom, I admire the way you manage all the different projects. I remember the first time we met you were pregnant, and that inspired me during my own pregnancy. You always found the time to help me with data interpretation, encouraged me with the most difficult findings, as with FLICA, and inspired me by proposing new ideas for different studies. I'm grateful for all of that and hope we can collaborate in the future.

I also want to thank Mark and Tony. It goes without saying that your help with data interpretation has been crucial for all my studies. I learned a lot from you on autism and infant research, but also on how to write an exciting paper. And Tony, I've always found very inspiring your talks about your career and the possibilities that we have as young researchers. Then of course I want to thank all participants and their families for their precious contribution, and researchers from the BASIS Team for collecting the data. I can only imagine the hard work necessary to put together such a great dataset. None of the studies reported in this thesis could have ever been performed without you. In particular, I want to thank Charlotte for collecting and analysing the ERP data. It has been great collaborating with you to investigate neural sensitivity to faces in infancy from different perspectives.

I must also thank my fellow PhD students, all the members of the BRAINVIEW network and the funders, the European Commission's Horizon 2020 Program. Being part of such a cross-disciplinary network in an internationally competitive scenario has been very motivating. It gave me the opportunity to meet and collaborate with amazing people, to travel all around Europe and to organize an international conference together with my fellow PhD students (ETADE). Special thanks go to the "London girls": Anna, Judit, Rianne and Tessel. You welcomed me in London for my placement and visits, and it has been great being on stage with you at INSAR 2018. I want to thank Anna for collaborating with me – I learned a lot on EEG microstates from that. And I want to thank Renata for always coming to some Asiatic restaurant with me during our training weeks and for giving me the idea for the cover. Special thanks go to her and Emma for being my paranymphs. But I also want to thank all the other PhD students for the great time spent together during our meetings every 6 months. Hope we will collaborate again in the future!
As I said, BRAINVIEW gave me the opportunity to travel a lot for collaborations. One placement I'll never forget is the one in Iceland and I want to thank all the people from Mentis Cura for that. In particular, I want to thank Ivar and Kristinn for welcoming me, teaching me a lot about the genetic algorithm and supervising me during my placement; Magdalena for teaching me a lot about EEG and for always having a big smile on her face; as well as Magnus, Bergrun, Peter, Herborg and Arni for the nice chats at lunchtime.

I'm at the end of this section and cannot forget the ones who really made all possible: my family. I want to thank my parents and my brother for supporting me over the years, always believing in me and encouraging me to achieve my goals. And you keep doing it now long-distance, as you let me go build my new life in the Netherlands. I'll never be grateful enough for all the sacrifices you have made for me! Then my husband, Ruben. Thank you for choosing to move to the Netherlands with me. Over the past three years, I grew a lot as a researcher but even more as a person. In fact, my whole life changed, starting from the country we live in, then our marriage and our first daughter. And none of this would have been possible without you by my side. You always say I'm the strong one, but I take all the strength I need from you. You have always supported me, encouraged me and believed in me, and I'll always be grateful to you for that! Finally, I want to thank my daughter, Jana, for inspiring me and giving me the strength to handle this thesis by the end of my pregnancy. I'll never forget the days and nights spent writing while feeling you from the inside. I dedicate this to you and hope it will inspire you and teach you that you need to work hard to do what you like.

ABOUT THE AUTHOR

Giorgia Bussu was born in Civitavecchia, Italy, on the 2nd of October 1990. In 2009, she graduated from high school and decided to start a bachelor in Physics at the University of Rome “La Sapienza”. In 2012, she graduated cum laude. Then, she continued her education with a master in Physics, where, given her interest in medical science, she specialized in Medical Physics. There she gained a particular interest in neuroscience and brain imaging. In line with these interests, she conducted for her master thesis an fMRI study to quantify the neuromodulatory effects of visual spatial attention on the visual network and the default-mode network in adults. She graduated cum laude in 2015. Following her graduation, she moved to the Netherlands to start a PhD at the Donders Institute on multi-modal multi-domain data integration and prediction of autism in infancy. There she was part of the Marie Sklodowska-Curie International Training Network “Brainview”, aimed in the development of novel methods to detect autism and other neurodevelopmental disorders at an early age. As a PhD candidate, she investigated developmental trajectories in infants at high and low familial risk for autism. In particular, she built integrated models on longitudinal data from different domains and different modalities for individual prediction of autism at an early age. Now she works as a postdoctoral researcher on imaging genetics at the Donders Institute under the AIMS-2-Trials project.

LIST OF PUBLICATIONS

- DiNuzzo, M., D. Mascali, M. Moraschi, G. Bussu, B. Maraviglia, S. Mangia and F. Giove (2017). “Temporal Information Entropy of the Blood-Oxygenation Level-Dependent Signals Increases in the Activated Human Primary Visual Cortex.” Front Phys **5**.

- Bussu, G., E. J. H. Jones, T. Charman, M. H. Johnson, J. K. Buitelaar andThe BASIS Team. (2018). “Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis.” J Autism Dev Disord.

- Pijl, M. K. J., G. Bussu, T. Charman, M. H. Johnson, E. J. H. Jones, G. Pasco, I. J. Oosterling, N. N. J. Rommelse, J. K. Buitelaar and The BASIS Team. (2019). “Temperament as an Early Risk Marker for Autism Spectrum Disorders? A Longitudinal Study of High-Risk and Low-Risk Infants.” J Autism Dev Disord.

- Bussu, G., Jones, E.J.H., Charman, T., Johnson, M.H., Buitelaar, J.K., and The BASIS Team. (2019) “Latent trajectories of adaptive behaviour in infants at high and low familial risk for autism spectrum disorder.” Mol Autism.

- Tye, C., Bussu, G., Jones, E.J.H., Gliga, T., Elsabbagh, M., Pasco, G., Johnsen, K., Charman, T., Buitelaar, J.K., Johnson, M.H., and the BASIS team. (in revision) “Widespread atypical neural responses to faces at 8 months predict autism spectrum at 3 years.”

- Bussu, G., Llera, A., Jones, E. J. H., Tye, C., Charman, T., Johnson, M. H., Beckmann, C.F., Buitelaar, J. K., & The BASIS Team. (in revision) “Uncovering Developmental Paths to Autism Spectrum Disorder through an Integrated Analysis of Developmental Measures and Neural Sensitivity to Faces.”

- Gui, A., Bussu, G., Tye, C., Elsabbagh, M., Pasco, G., Charman, T., Johnson, M. H., and Jones, E. J. H. (in revision) “Diminished engagement of attentive brain states to faces predicts later autism.”

DONDERS GRADUATE SCHOOL FOR COGNITIVE NEUROSCIENCE SERIES

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:
http://www.ru.nl/donders/graduate-school/phd/


DONDERS
INSTITUTE


Radboud University
IN DEI NOMINE FELICITER
Radboudumc